

Achieving the Secrecy Capacity of Wiretap Channels Using Polar Codes

Hessam Mahdavifar, *Student Member, IEEE*, and Alexander Vardy, *Fellow, IEEE*

Abstract—Suppose that Alice wishes to send messages to Bob through a communication channel C_1 , but her transmissions also reach an eavesdropper Eve through another channel C_2 . This is the wiretap channel model introduced by Wyner in 1975. The goal is to design a coding scheme that makes it possible for Alice to communicate both *reliably* and *securely*. Reliability is measured in terms of Bob's probability of error in recovering the message, while security is measured in terms of the mutual information between the message and Eve's observations. Wyner showed that the situation is characterized by a single constant C_s , called the *secrecy capacity*, which has the following meaning: for all $\varepsilon > 0$, there exist coding schemes of rate $R \geq C_s - \varepsilon$ that asymptotically achieve both the reliability and the security objectives. However, his proof of this result is based upon a nonconstructive random-coding argument. To date, despite a considerable research effort, the only case where we know how to *construct* coding schemes that achieve secrecy capacity is when Eve's channel C_2 is an erasure channel, or a combinatorial variation thereof.

Polar codes were recently invented by Arıkan; they approach the capacity of symmetric binary-input discrete memoryless channels with low encoding and decoding complexity. In this paper, we use polar codes to construct a coding scheme that achieves the secrecy capacity for a wide range of wiretap channels. Our construction works for any instantiation of the wiretap channel model, as long as both C_1 and C_2 are symmetric and binary-input, and C_2 is degraded with respect to C_1 . Moreover, we show how to modify our construction in order to provide *strong security*, in the sense defined by Maurer, while still operating at a rate that approaches the secrecy capacity. In this case, we cannot guarantee that the reliability condition will also be satisfied unless the main channel C_1 is noiseless, although we believe it can be always satisfied in practice.

Index Terms—channel polarization, information-theoretic security, polar codes, secrecy capacity, strong security, wiretap channel

I. INTRODUCTION

THE notion of wiretap channels was introduced by Aaron Wyner [42] in 1975. In this setting, Alice wishes to send messages to Bob through a communication channel C_1 , called the *main channel*, but her transmissions also reach an adversary Eve through another channel C_2 , called the *wiretap channel*. This is illustrated in Figure 1, wherein \mathbf{U} denotes a k -bit message that Alice wishes to communicate to Bob. We think of \mathbf{U} as a random variable that takes values in $\{0, 1\}^k$; unlike most papers on wiretap channels, we *do not assume anything* regarding the a priori distribution of \mathbf{U} . While making use of auxiliary random bits, the encoder maps \mathbf{U} into a sequence \mathbf{X} of n chan-

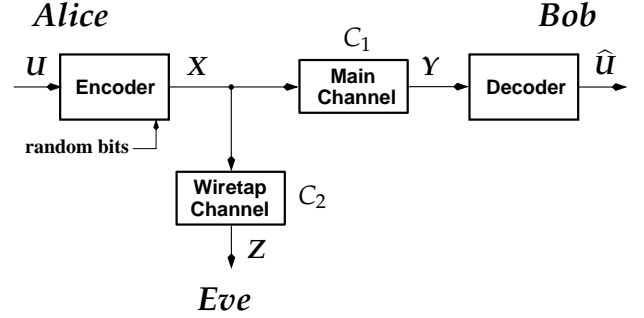


Figure 1. Block diagram of a generic wiretap-channel system

nel symbols. This sequence is transmitted across the main channel and the wiretap channel resulting in the corresponding channel outputs \mathbf{Y} and \mathbf{Z} . Finally, the decoder maps \mathbf{Y} (deterministically) into an estimate $\hat{\mathbf{U}}$ of the original message.

The goal is to design a coding scheme — namely, an encoding algorithm and a decoding algorithm — that makes it possible to communicate both *reliably* and *securely*, as the message length k tends to infinity. Reliability is measured in terms of the *probability of error* in recovering the message. Specifically, the objective is to satisfy the following

$$\text{Reliability Condition: } \lim_{k \rightarrow \infty} \Pr\{\hat{\mathbf{U}} \neq \mathbf{U}\} = 0 \quad (1)$$

where the probability is over all the relevant coin tosses in the system: in the generation of \mathbf{U} , in the encoder, and in the main channel. Security is usually measured in terms of the *normalized mutual information* between the message \mathbf{U} and Eve's observations \mathbf{Z} . Specifically, one is interested in encoding algorithms that satisfy the following

$$\text{Security Condition: } \lim_{k \rightarrow \infty} \frac{I(\mathbf{U}; \mathbf{Z})}{k} = 0 \quad (2)$$

Note that $I(\mathbf{U}; \mathbf{Z})$ is equal to the difference between the a priori entropy $H(\mathbf{U})$ and the conditional entropy $H(\mathbf{U}|\mathbf{Z})$. Thus, intuitively, (2) means that observing \mathbf{Z} does not provide much information about \mathbf{U} beyond what is available a priori, as compared to the message length k . Maurer argued in [28, 29] that the conventional notion of security (2) is much too weak. Indeed, it is easy to construct examples where $k^{1-\varepsilon}$ out of the k message bits are disclosed to Eve, while still satisfying (2). This is clearly unacceptable. Thus Maurer introduced in [28] an alternative

$$\text{Strong Security Condition: } \lim_{k \rightarrow \infty} I(\mathbf{U}; \mathbf{Z}) = 0 \quad (3)$$

Notice that both security conditions (2) and (3) are information-theoretic rather than computational: the adversary is assumed to be computationally unbounded, and security does not depend on computational hardness assumptions of any kind.

The material in this paper was presented in part at the IEEE International Symposium on Information Theory, Austin, Texas, June 2010.

Hessam Mahdavifar is with the Department of Electrical and Computer Engineering, University of California San Diego, La Jolla, CA 92093-0407, U.S.A. (e-mail: hessam@ucsd.edu).

Alexander Vardy is with the Department of Electrical and Computer Engineering, the Department of Computer Science and Engineering, and the Department of Mathematics, University of California San Diego, La Jolla, CA 92093-0407, U.S.A. (e-mail: avardy@ucsd.edu).

A. Prior Work

In 1975, Wyner [42] considered a special case of the system in Figure 1 where both C_1 and C_2 are discrete memoryless channels (DMCs) and, moreover, C_2 is degraded with respect to C_1 . He proved that such a system is characterized by a single constant C_s , called the *secrecy capacity*, which has the following meaning. For all $\varepsilon > 0$, there exist coding schemes of information rate $R \geq C_s - \varepsilon$ that satisfy (1) and (2); conversely, it is not possible to satisfy both (1) and (2) at rates greater than C_s . Since 1975, Wyner's results have been extended to a variety of contexts, most notably Gaussian channels [23], general broadcast channels with confidential messages [12], and channels that impose a combinatorial (rather than probabilistic) constraint on the adversary [8, 33]. In fact, the literature on wiretap channels encompasses, by now, hundreds of papers.

However, the vast majority of this work relies on nonconstructive random-coding arguments to establish the main results. Such results show that *there exist* codes that achieve secrecy capacity, but are of little use if one's goal is to design specific polynomial-time encoding/decoding algorithms. To the best of our knowledge, *constructive solutions* to the wiretap-channel problem are available only in two special cases. The first special case is when the main channel is noiseless and the wiretap channel is the binary erasure channel (BEC). A coding scheme for this case, using LDPC codes for the BEC, was presented in [37, 39] and proved to achieve secrecy capacity. The other special case is when the adversary is constrained combinatorially: Eve can select to observe some t out of the n transmitted symbols, while the remaining $n - t$ symbols are erased. This situation, studied by Ozarow and Wyner in [33], may be regarded as a combinatorial variation of an erasure channel. Provably optimal coding schemes for this case can be constructed from MDS codes [41], or using extractors [8]. We observe, however, that even for the simple situation where C_1 is noiseless and C_2 is a binary symmetric channel, it is not known how to explicitly construct codes that achieve secrecy capacity.

We point out that a general method of coding for the wiretap channel, often referred to as *coset-coding* or *syndrome-coding*, is well known. This method goes back to the work of Wyner [33, 42], although it was significantly extended and generalized in [9, 10] and other papers. Assume, for simplicity, that the input alphabet of both C_1 and C_2 is binary. In this case, the coset-coding method utilizes two binary linear codes: an “outer” code C^* and an “inner” code C , such that $C \subset C^*$ and the difference $\dim(C^*) - \dim(C)$ between their dimensions is k . This condition implies that C^* can be partitioned into 2^k cosets of C . A message $u \in \{0, 1\}^k$ is conveyed by Alice via the choice of one of these 2^k cosets, say $a + C$. What is transmitted by Alice is a vector X that is selected uniformly at random from $a + C$. Loosely speaking, the outer code C^* serves to correct the errors on the main channel, and thus ensures reliability, while the inner code C , over which X is randomized, ensures security. The trouble is that it is not known how to *explicitly construct* a sequence of outer codes C^* and inner codes C that satisfy conditions (1) and (2) at a rate that approaches the secrecy capacity as $n \rightarrow \infty$. The work of Cohen and Zémor [9, 10] shows that a *random choice* of the inner code C suffices to achieve

strong security, in a very general setting. Notably, the proof of this result in [9, 10] does *not* assume that the messages are uniformly random a priori. Still, to the best of our knowledge, the only cases where explicit constructions of C and C^* are known are those described in the foregoing paragraph (cf. [37, 39]).

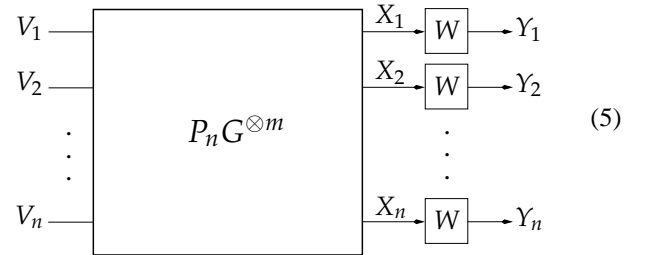
B. Our Contributions

In this paper, we present a coding scheme that achieves the secrecy capacity of wiretap channels whenever C_1 and C_2 are binary-input symmetric DMCs and C_2 is degraded with respect to C_1 . This is the situation originally studied by Wyner; it includes the important special case where C_1 and C_2 are arbitrary binary symmetric channels. We are able to satisfy the reliability and security conditions (1) and (2) with explicit polynomial-time encoding/decoding algorithms. In fact, the number of operations required for encoding and decoding is only $O(n \log n)$. Our construction is based upon key results in the literature on *polar codes*, recently invented by Arıkan [3].

It is proved in [3] that polar codes achieve the capacity of arbitrary binary-input symmetric DMCs, with low encoding and decoding complexity. The proof of this result is based on a phenomenon called *channel polarization*. Let

$$G = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \quad (4)$$

and let $G^{\otimes m}$ denote the m -th Kronecker power of G . Let W be a symmetric binary-input DMC, and let $V = (V_1, V_2, \dots, V_n)$ be a block of $n = 2^m$ bits chosen uniformly at random from $\{0, 1\}^n$. Suppose V is encoded as $X = VP_n G^{\otimes m}$, where P_n is the $n \times n$ bit-reversal permutation matrix. Finally, X is transmitted through n independent copies of W , as shown below:



Arıkan [3] considered the n channels “seen” by each of the n individual bits V_1, V_2, \dots, V_n as they undergo the transformation in (5). Let us call them the *bit-channels* — for a precise definition of the notion of a bit-channel, see [3] and Section III. It is shown in [3] that as m grows, the bit-channels start polarizing: they approach either a noiseless channel or a pure-noise channel. We will say that the former bit-channels are *good* while the latter are *bad* (again, see Section III for a rigorous definition). One of the key results of [3] is that the fraction of bit-channels that are good approaches the capacity of W as $n \rightarrow \infty$.

Given the channel polarization phenomenon, the general idea of our construction is quite simple. We will transmit random bits over those bit-channels that are good for both Eve and Bob, information bits over those bit-channels that are good for Bob but bad for Eve, and zeros over those bit-channels that are bad for both Bob and Eve. In the rest of this paper, we make this idea precise and prove that it works.

In Section II, we briefly recap relevant results from the literature on wiretap channels, in order to obtain a simple expression for the secrecy capacity C_s in the case where C_1 and C_2 are symmetric DMCs and C_2 is degraded with respect to C_1 . In Section III, we provide the necessary background on polar codes and establish a certain property of channel polarization that is crucial for our construction (Lemma 4). The construction itself, namely the proposed coding scheme, is presented in Section IV. In Section V, we prove that the proposed coding scheme satisfies the reliability and security conditions (1) and (2). We also show in Section V that the rate k/n of our coding scheme approaches the secrecy capacity C_s as $n \rightarrow \infty$.

In Section VI, we consider the stronger notion of security (3). It was shown by Maurer-Wolf [29] that any coding scheme that satisfies the “weak” security condition (2) can be converted into a coding scheme that satisfies the stronger condition (3). This is accomplished using an ingenious information reconciliation and privacy amplification protocol [7]. Although, in principle, the rate overhead necessary for privacy amplification can be made arbitrarily small, this is unlikely to be the case in practice. In Section VI, we show how to modify the coding scheme of Section IV in order to guarantee strong security *directly*, without the need for privacy amplification. This is achieved by suitably modifying our definition of “bad” bit-channels, in a manner that differs from the generally accepted notions [3]. As a result, under the modified definition, a vanishing fraction of bit-channels could be good for Eve but bad for Bob, even when the wiretap channel is degraded with respect to the main channel. In this situation, the rate of the coding scheme of Section VI still approaches the secrecy capacity C_s , but we cannot guarantee that the reliability condition (1) is satisfied, unless the main channel is noiseless. Nevertheless, we believe that, in practice, acceptably low probabilities of error could be achieved on the main channel (using a more elaborate decoding algorithm).

We conclude the paper in Section VII with a brief discussion of further results. In particular, we explain in Section VII that our construction generalizes straightforwardly to the case where the channels C_1 and C_2 are *not* symmetric, although in this case polar codes become less explicit and only the “symmetric secrecy capacity” can be achieved. A few open problems that stem from our results herein are also discussed in Section VII.

C. Related Work

Following the publication of a preliminary version of this paper in [25] and [26], several related papers have appeared [1, 17, 21]. Most notably, the work of Hof and Shamai [17] on polar coding for wiretap channels is independent and contemporaneous to ours. While some of the main results in [17] and in this paper are similar, there are important differences that we would like to emphasize. One key difference is that Hof and Shamai [17] analyze their polar-coding scheme in detail, including recursive channel combining and splitting, whereas we treat polar codes essentially as a black box. We believe this makes our proof both shorter and clearer.

There are also significant differences between the results established in [1, 17, 21] and in this paper. In particular, it is shown in [1, 17] that polar coding achieves the entire *rate-equivocation*

region (see [24] for a definition), whereas we are interested only in the extreme point of this region that corresponds to secrecy capacity. On the other hand, in several other respects, our results are stronger than those of [1, 17, 21]. First, the proof in [1, 17] is contingent on the assumption that the message \mathbf{U} is a priori uniform over $\{0, 1\}^k$, whereas we do not place any constraints on the a priori distribution of \mathbf{U} . Assuming that messages are a priori uniform is common in information theory, but such assumptions are completely unacceptable in cryptography [6, 15]. Even more importantly, we show how polar coding should be used to provide *strong security*, whereas the work of [1, 17, 21] provides weak security only. Again, in cryptographic applications, conventional weak security is usually unacceptable.

II. SECRECY CAPACITY

In this section, we first establish some relevant terminology. We then briefly recap the results of [12, 22] to provide a simple expression for the secrecy capacity C_s in the case where C_1 and C_2 are symmetric DMCs and C_2 is degraded with respect to C_1 .

We will limit our consideration to finite-input and finite-output discrete memoryless channels throughout. Such a **channel** is a triple $\langle \mathcal{X}, \mathcal{Y}, W \rangle$, where \mathcal{X}, \mathcal{Y} are finite sets and W is an $|\mathcal{X}| \times |\mathcal{Y}|$ matrix with $W[x, y]$ being the probability of receiving $y \in \mathcal{Y}$ given that $x \in \mathcal{X}$ was sent. We will follow the convention of [3, 4, 19] and write $W(y|x)$ instead of $W[x, y]$.

A matrix M is strongly symmetric if the rows of M are permutations of each other and the columns of M are permutations of each other. A channel $\langle \mathcal{X}, \mathcal{Y}, W \rangle$ is **strongly symmetric** if W is a strongly symmetric matrix. Following [3, 4, 14, 19], we will say that $\langle \mathcal{X}, \mathcal{Y}, W \rangle$ is **symmetric** (often called *output-symmetric*) if the columns of W can be partitioned into subsets such that each subset forms a strongly symmetric matrix. The **capacity** of a symmetric channel $\langle \mathcal{X}, \mathcal{Y}, W \rangle$ is given by

$$\mathcal{C}(W) \stackrel{\text{def}}{=} H(X) - H(X|Y) = \log_2 |\mathcal{X}| - H(X|Y) \quad (6)$$

where the random variable X at the input to the channel is uniform over \mathcal{X} , and Y is the corresponding random variable at the channel output (for a proof of this fact, see [14, p. 94]). An important example of a symmetric channel is the **binary symmetric channel** $\text{BSC}(p) = \langle \{0, 1\}, \{0, 1\}, W \rangle$ with

$$W = \begin{bmatrix} 1-p & p \\ p & 1-p \end{bmatrix}$$

Given a channel $C_1 = \langle \mathcal{X}, \mathcal{Y}, W_1 \rangle$, we say that another channel $C_2 = \langle \mathcal{X}, \mathcal{Z}, W_2 \rangle$ is **degraded with respect to** C_1 if there exists a third channel $C_3 = \langle \mathcal{Y}, \mathcal{Z}, W_3 \rangle$ such that C_2 is the cascade of C_1 and C_3 . Specifically, it is required that

$$W_2(z|x) = \sum_{y \in \mathcal{Y}} W_1(y|x) W_3(z|y) \quad (7)$$

for all $x \in \mathcal{X}$ and $z \in \mathcal{Z}$. Note that whenever $p_2 \geq p_1$, the channel $C_2 = \text{BSC}(p_2)$ is degraded with respect to $C_1 = \text{BSC}(p_1)$.

The **secrecy capacity** C_s of the wiretap-channel system in Figure 1 is defined as follows. First, assume that the message \mathbf{U} is uniformly random over $\{0, 1\}^k$. Then C_s is the supremum over all rates $R = k/n$ (in bits per channel use) such that there exist coding schemes of rate R satisfying conditions (1) and (2). For the general case where C_1 and C_2 are arbitrary DMCs, comput-

ing the secrecy capacity is a difficult problem. Let X denote the single-letter input to C_1 and C_2 , let Y and Z denote the corresponding single-letter outputs. The best known expression for the secrecy capacity C_s , given by Csiszár and Körner in [12], is

$$C_s = \max_U \left(I(U; Y) - I(U; Z) \right)$$

where the maximum is taken over all random variables U such that $U \rightarrow X \rightarrow (Y, Z)$ is a Markov chain. The problem is that this maximization is often difficult to evaluate, and there is no simpler expression for the secrecy capacity even when C_1 and C_2 are both strongly symmetric, unless additional constraints are satisfied. See [24, 40] for more details on this.

However, when $C_1 = \langle \mathcal{X}, \mathcal{Y}, W^* \rangle$ and $C_2 = \langle \mathcal{X}, \mathcal{Z}, W \rangle$ are symmetric and C_2 is degraded with respect to C_1 , a simple expression for C_s was given by Leung-Yan-Cheong in [22]. It is shown in [22, Theorem 4] that in this case

$$C_s = \mathcal{C}(W^*) - \mathcal{C}(W) = H(X|Z) - H(X|Y) \quad (8)$$

where X is uniform over \mathcal{X} . In particular, if the main channel is $\text{BSC}(p_1)$ while the wiretap channel is $\text{BSC}(p_2)$, with $p_2 \geq p_1$, then the secrecy capacity is given by $h_2(p_2) - h_2(p_1)$, where $h_2(\cdot)$ is the binary entropy function.

III. POLAR CODES

This section provides a concise overview of the groundbreaking work of Arıkan [3] and others [4, 19, 20] on polar codes and channel polarization. We establish only those results that are essential for the coding schemes presented in this paper.

As in [16, 20], we consider exclusively **binary-input symmetric memoryless** (BSM) discrete channels. Such a channel is a symmetric DMC, as defined in the previous section, with input alphabet $\mathcal{X} = \{0, 1\}$. With a slight abuse of notation, we will often follow [3, 4, 20] and simply write W to denote a BSM channel $\langle \{0, 1\}, \mathcal{Y}, W \rangle$. The **Bhattacharyya parameter** of W is

$$Z(W) \stackrel{\text{def}}{=} \sum_{y \in \mathcal{Y}} \sqrt{W(y|0)W(y|1)}$$

It can be shown that $Z(W)$ always takes values in $[0, 1]$. Intuitively, channels with $Z(W) \leq \varepsilon$ are almost noiseless, while channels with $Z(W) \geq 1 - \varepsilon$ are almost pure-noise channels. This intuition is made precise in [3, Proposition 1].

Arıkan [3] introduces a number of channels that are associated with the transformation in (5). First, there is the channel $\langle \{0, 1\}^n, \mathcal{Y}^n, W^n \rangle$ given by

$$W^n(\mathbf{y}|\mathbf{x}) \stackrel{\text{def}}{=} \prod_{i=1}^n W(y_i|x_i) \quad (9)$$

where $\mathbf{x} = (x_1, x_2, \dots, x_n)$ and $\mathbf{y} = (y_1, y_2, \dots, y_n)$. This is the channel that results from n independent uses of the channel W . Next, for all $n = 2^m$, let us define the **Arıkan transform matrix** $G_n \stackrel{\text{def}}{=} P_n G^{\otimes m}$, where G is the matrix in (4) and P_n is the bit-reversal permutation matrix defined in [3, Section VII-B]. Arıkan then introduces the “combined” channel $\langle \{0, 1\}^n, \mathcal{Y}^n, \tilde{W} \rangle$ with transition probabilities given by

$$\tilde{W}(\mathbf{y}|\mathbf{v}) \stackrel{\text{def}}{=} W^n(\mathbf{y}|\mathbf{v} G_n) = W^n(\mathbf{y}|\mathbf{v} P_n G^{\otimes m}) \quad (10)$$

This is the channel seen by the random vector (V_1, V_2, \dots, V_n) as it undergoes the transformation in (5). Arıkan [3] also defines

the channel $\langle \{0, 1\}, \mathcal{Y}^n \times \{0, 1\}^{i-1}, W_i \rangle$ that is seen by the i -th bit V_i , for $i = 1, 2, \dots, n$, as follows. Let $\mathbf{v}_i = (v_1, v_2, \dots, v_i)$ denote a binary vector of length i , with the convention that \mathbf{v}_0 is the empty string and that $\{0, 1\}^0 = \{\mathbf{v}_0\}$. Then

$$W_i(\mathbf{y}, \mathbf{v}_{i-1}|\mathbf{v}_i) \stackrel{\text{def}}{=} \frac{1}{2^{n-1}} \sum_{\bar{\mathbf{v}} \in \{0, 1\}^{n-i}} \tilde{W}(\mathbf{y}|\mathbf{v}_{i-1}, \mathbf{v}_i, \bar{\mathbf{v}}) \quad (11)$$

where (\cdot, \cdot) denotes vector concatenation. It is easy to show (cf. Lemma 14) that $W_i(\mathbf{y}, \mathbf{v}_{i-1}|\mathbf{v}_i)$ is indeed the probability of the event that $(Y_1, Y_2, \dots, Y_n) = \mathbf{y}$ and $(V_1, V_2, \dots, V_{i-1}) = \mathbf{v}_{i-1}$ given the event $V_i = v_i$, provided $\mathbf{V} = (V_1, V_2, \dots, V_n)$ is a priori uniform over $\{0, 1\}^n$. Consequently, if one considers a “hypothetical decoder” that attempts to estimate the i -th bit V_i having observed \mathbf{y} and \mathbf{v}_{i-1} , then W_i is the effective channel seen by such decoder (again, provided \mathbf{V} is a priori uniform). We will refer to W_i as the **i -th bit-channel**, for $i = 1, 2, \dots, n$.

Observe that the optimal decision rule for the hypothetical decoder of the foregoing paragraph is trivial: decide $\hat{v}_i = 0$ if

$$W_i(\mathbf{y}, \mathbf{v}_{i-1}|0) \geq W_i(\mathbf{y}, \mathbf{v}_{i-1}|1) \quad (12)$$

and $\hat{v}_i = 1$ otherwise. One can invoke this decision rule iteratively for all $i = 1, 2, \dots, n$, while substituting the first $i - 1$ decisions $(\hat{v}_1, \hat{v}_2, \dots, \hat{v}_{i-1})$ in place of the hypothetical observations \mathbf{v}_{i-1} . Up to a small modification described later, this is the **successive cancellation decoder** invented by Arıkan [3].

Following [4, 19], let us partition the n bit-channels into **good channels** and **bad channels** as follows. Let $[n] \stackrel{\text{def}}{=} \{1, 2, \dots, n\}$ and let $\beta < 1/2$ be a fixed positive constant. Then the index sets of the good and bad channels are given by

$$\mathcal{G}_n(W, \beta) \stackrel{\text{def}}{=} \{i \in [n] : Z(W_i) < 2^{-n^\beta/n}\} \quad (13)$$

$$\mathcal{B}_n(W, \beta) \stackrel{\text{def}}{=} \{i \in [n] : Z(W_i) \geq 2^{-n^\beta/n}\} \quad (14)$$

One of the key results of [3, 4] is that the fraction of the good channels approaches the channel capacity $\mathcal{C}(W)$, given by (6), as $n \rightarrow \infty$. We state this result precisely as follows.

Theorem 1. *For any BSM channel W and any constant $\beta < 1/2$ we have*

$$\lim_{n \rightarrow \infty} \frac{|\mathcal{G}_n(W, \beta)|}{n} = \mathcal{C}(W)$$

Theorem 1 readily leads to a construction of capacity-achieving **polar codes**. The general idea is to transmit the information bits over the good bit-channels while fixing the input to the bad bit-channels to a priori known values, say zeros¹. Formally, given a vector \mathbf{v} of length n and a set $\mathcal{A} \subseteq [n]$, let $\mathbf{v}_{\mathcal{A}}$ denote the projection of \mathbf{v} on the coordinates in \mathcal{A} . Each subset \mathcal{A} of $[n]$ of size $|\mathcal{A}| = k$ specifies a **polar code** $\mathbf{C}_n(\mathcal{A})$ of rate k/n . We define $\mathbf{C}_n(\mathcal{A})$ via its encoder map $\mathcal{E} : \{0, 1\}^k \rightarrow \{0, 1\}^n$. Given a message $\mathbf{u} \in \{0, 1\}^k$, the encoder proceeds in two steps. First, the encoder constructs the vector $\mathbf{v} \in \{0, 1\}^n$, by setting $\mathbf{v}_{\mathcal{A}} = \mathbf{u}$ and $\mathbf{v}_{\mathcal{A}^c} = \mathbf{0}$, where \mathcal{A}^c is the complement of \mathcal{A} in $[n]$ and $\mathbf{0}$ is

¹In the work of [3, 4], which deals with symmetric as well as non-symmetric DMCs, it is important to allow an arbitrary choice of these “frozen” values. However, it is also shown in [3] that in the case of symmetric channels, any choice is as good as any other. Since we are concerned exclusively with symmetric channels in this paper, we will use zeros for notational convenience.

the all-zero vector. Next, it outputs $\mathcal{E}(\mathbf{u}) = \mathbf{v} G_n$ as in (5). The decoder we will use for $\mathbb{C}_n(\mathcal{A})$ is the successive cancellation decoder of Arkan [3]. This decoder works as already described in (12), with one straightforward modification: for $i \in \mathcal{A}^c$, the decision rule is simply $\hat{v}_i = 0$.

The key property of the encoder-decoder pair of the foregoing paragraph is summarized in the following theorem. This theorem is (the second part of) Proposition 2 of Arkan [3].

Theorem 2. *Let W be a BSM channel and let \mathcal{A} be an arbitrary subset of $[n]$ of size $|\mathcal{A}| = k$. Suppose that a message \mathbf{U} is chosen uniformly at random from $\{0, 1\}^k$, encoded as a codeword of $\mathbb{C}_n(\mathcal{A})$, and transmitted over W . Then the probability that the channel output is not decoded to \mathbf{U} under successive cancellation decoding satisfies*

$$\Pr\{\hat{\mathbf{U}} \neq \mathbf{U}\} \leq \sum_{i \in \mathcal{A}} Z(W_i) \quad (15)$$

In this paper, we need a result that is somewhat stronger than Theorem 2, since we do *not* assume that messages are chosen uniformly at random from $\{0, 1\}^k$. Fortunately, such a result can be readily established using the machinery that was already developed by Arkan [3] for symmetric channels. Indeed, for symmetric channels, it is well known that the error probability is independent of the transmitted codeword; hence, the input distribution should not matter. The following proposition makes this observation precise. We include a proof, for completeness.

Proposition 3. *Let W be a BSM channel and let \mathcal{A} be an arbitrary subset of $[n]$ of size $|\mathcal{A}| = k$. Suppose that a message \mathbf{U} is chosen according to an arbitrary distribution from $\{0, 1\}^k$, encoded as a codeword of $\mathbb{C}_n(\mathcal{A})$, and transmitted over W . Then the probability P_e that the channel output is not decoded to \mathbf{U} under successive cancellation decoding satisfies*

$$P_e \leq \sum_{i \in \mathcal{A}} Z(W_i) \quad (16)$$

Proof. Following Arkan [3, Section V], we consider the sample space $\Omega_n = \{0, 1\}^n \times \mathcal{Y}^n$ with the probability measure

$$\Pr\{(\mathbf{v}, \mathbf{y})\} \stackrel{\text{def}}{=} 2^{-n} \tilde{W}(\mathbf{y}|\mathbf{v}) \quad (17)$$

for all $\mathbf{v} \in \{0, 1\}^n$ and $\mathbf{y} \in \mathcal{Y}^n$. On this probability space, Arkan [3] defines the event \mathcal{E} of block error as the set of all pairs (\mathbf{v}, \mathbf{y}) in Ω_n such that the channel output \mathbf{y} is not decoded to $\mathbf{v}_\mathcal{A}$ under successive cancellation decoding. Let us further define, for all $\mathbf{w} \in \{0, 1\}^n$, the event

$$\mathcal{V}_\mathbf{w} \stackrel{\text{def}}{=} \{(\mathbf{v}, \mathbf{y}) \in \Omega_n : \mathbf{v} = \mathbf{w}\} \quad (18)$$

It is shown in [3, Proposition 2] that under the probability measure in (17) we have

$$\Pr\{\mathcal{E}\} \leq \sum_{i \in \mathcal{A}} Z(W_i) \quad (19)$$

It is furthermore shown in [3, Section VI-B] that, provided ties in the decision rule (12) are broken at random, the events \mathcal{E} and $\mathcal{V}_\mathbf{w}$ are independent for all \mathbf{w} . In other words,

$$\Pr\{\mathcal{E} | \mathcal{V}_\mathbf{w}\} = \Pr\{\mathcal{E}\} \quad \text{for all } \mathbf{w} \in \{0, 1\}^n \quad (20)$$

Now consider the situation where the a priori distribution on the messages in $\{0, 1\}^k$ is a delta-function. That is, a specific mes-

sage $\mathbf{u} \in \{0, 1\}^k$ is always chosen with probability 1, and the input to the transformation in (5) is the vector \mathbf{w} with $\mathbf{w}_\mathcal{A} = \mathbf{u}$ and $\mathbf{w}_{\mathcal{A}^c} = \mathbf{0}$. Let $P_e(\mathbf{u})$ denote the probability that the successive cancellation decoder does not decode the corresponding channel output \mathbf{y} to $\mathbf{w}_\mathcal{A}$. Then it follows from (10) that

$$P_e(\mathbf{u}) = \sum_{\mathbf{y} \in \mathcal{F}(\mathbf{w}_\mathcal{A})} \tilde{W}(\mathbf{y}|\mathbf{w}) \quad (21)$$

where $\mathcal{F}(\mathbf{w}_\mathcal{A})$ is the set of channel outputs that are not decoded to $\mathbf{w}_\mathcal{A}$ by the successive cancellation decoder. Observe that

$$\mathcal{E} \cap \mathcal{V}_\mathbf{w} = \{(\mathbf{v}, \mathbf{y}) \in \Omega_n : \mathbf{v} = \mathbf{w} \text{ and } \mathbf{y} \in \mathcal{F}(\mathbf{w}_\mathcal{A})\}$$

Therefore, the probability of the event $\mathcal{E} \cap \mathcal{V}_\mathbf{w}$, under the probability measure in (17), can be expressed as

$$\Pr\{\mathcal{E} \cap \mathcal{V}_\mathbf{w}\} = \sum_{\mathbf{y} \in \mathcal{E} \cap \mathcal{V}_\mathbf{w}} 2^{-n} \tilde{W}(\mathbf{y}|\mathbf{v}) = 2^{-n} \sum_{\mathbf{y} \in \mathcal{F}(\mathbf{w}_\mathcal{A})} \tilde{W}(\mathbf{y}|\mathbf{w}) \quad (22)$$

It can be readily seen from (18) that $\Pr\{\mathcal{V}_\mathbf{w}\} = 2^{-n}$ for all \mathbf{w} . Hence, it follows from (20) that

$$\Pr\{\mathcal{E} \cap \mathcal{V}_\mathbf{w}\} = \Pr\{\mathcal{E} | \mathcal{V}_\mathbf{w}\} \Pr\{\mathcal{V}_\mathbf{w}\} = 2^{-n} \Pr\{\mathcal{E}\} \quad (23)$$

Combining (21), (22), (23), we conclude that $P_e(\mathbf{u}) = \Pr\{\mathcal{E}\}$. Since this holds for all $\mathbf{u} \in \{0, 1\}^k$, we have

$$P_e = \sum_{\mathbf{u} \in \{0, 1\}^k} P_e(\mathbf{u}) \Pr\{\mathbf{U} = \mathbf{u}\} = \Pr\{\mathcal{E}\}$$

for any probability distribution $\Pr\{\mathbf{U} = \mathbf{u}\}$ on $\{0, 1\}^k$. Hence, the proposition now follows from (19). ■

In order to establish our main result in Section V, we also need to consider a slightly different encoding and decoding scenario. In this variation, the encoder \mathcal{E}' is no longer deterministic. Rather, it has access to random bits and selects $\mathbf{V}_{\mathcal{A}^c}$ at random, according to some fixed (but otherwise arbitrary) probability distribution on $\{0, 1\}^{n-k}$. In all other respects, \mathcal{E}' is identical to the encoder \mathcal{E} for $\mathbb{C}_n(\mathcal{A})$ described above. The specific realization $\mathbf{v}_{\mathcal{A}^c}$ of $\mathbf{V}_{\mathcal{A}^c}$ is revealed to the decoder by a genie. The decoder uses successive cancellation, with the suitable modification: for $i \in \mathcal{A}^c$, the value of \hat{v}_i is not set to zero, but rather to the corresponding coordinate in the realization $\mathbf{v}_{\mathcal{A}^c}$ (revealed by the genie). Let P'_e denote the probability of block error in this scenario. It is shown in [3, Section VI] that Theorem 2 still applies in this case, namely

$$P'_e \leq \sum_{i \in \mathcal{A}} Z(W_i) \quad (24)$$

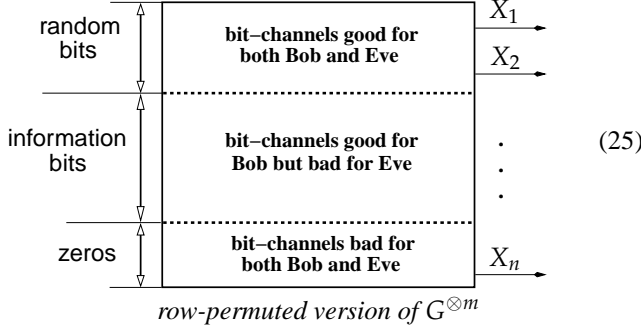
The coding scheme presented in the next section relies crucially on one more result from the literature on polar codes. The following lemma was proved by Korada in [19, Lemma 4.7].

Lemma 4. *Let W and W^* be BSM channels such that W is degraded with respect to W^* . For $n = 2^m$, let W_1, W_2, \dots, W_n and $W_1^*, W_2^*, \dots, W_n^*$ denote the n corresponding bit-channels. Then W_i is degraded with respect to W_i^* for all $i = 1, 2, \dots, n$, and therefore $\mathcal{C}(W_i) \leq \mathcal{C}(W_i^*)$ and $Z(W_i) \geq Z(W_i^*)$.*

It follows immediately from Lemma 4 and (13) that if W is degraded with respect to W^* , then the set of good bit-channels for W is a subset of the set of good bit-channels for W^* . More precisely, we have $\mathcal{G}_n(W, \beta) \subseteq \mathcal{G}_n(W^*, \beta)$ for all constants β .

IV. THE CODING SCHEME

We consider a special case of the wiretap-channel system of Figure 1, wherein both the main channel $C_1 = \langle \{0, 1\}, \mathcal{Y}, W^* \rangle$ and Eve's wiretap channel $C_2 = \langle \{0, 1\}, \mathcal{Z}, W \rangle$ are symmetric DMCs, and C_2 is degraded with respect to C_1 . The proposed coding scheme is illustrated informally below:



The general idea is to transmit information *only* over those bit-channels that are bad for Eve, while flooding those bit channels that are good for Eve with random bits. Formally, we fix a positive constant $\beta < 1/2$ and define three subsets of $[n]$ as follows:

$$\mathcal{R} \stackrel{\text{def}}{=} \mathcal{G}_n(W, \beta) \quad (26)$$

$$\mathcal{A} \stackrel{\text{def}}{=} \mathcal{G}_n(W^*, \beta) \setminus \mathcal{G}_n(W, \beta) \quad (27)$$

$$\mathcal{B} \stackrel{\text{def}}{=} \mathcal{B}_n(W^*, \beta) \quad (28)$$

Notice that the sets $\mathcal{R}, \mathcal{A}, \mathcal{B}$ are disjoint and $\mathcal{R} \cup \mathcal{A} \cup \mathcal{B} = [n]$. This is so since $\mathcal{G}_n(W^*, \beta)$ and $\mathcal{B}_n(W^*, \beta)$ are complements of each other by definition, and $\mathcal{G}_n(W, \beta) \subseteq \mathcal{G}_n(W^*, \beta)$ by Lemma 4. Let $|\mathcal{R}| = r$ and $|\mathcal{A}| = k$. We are now ready to describe the proposed encoding and decoding algorithms.

Encoding Algorithm: Formally, the encoder is a function $\mathcal{E}: \{0, 1\}^k \times \{0, 1\}^r \rightarrow \{0, 1\}^n$. It accepts as input a message $\mathbf{u} \in \{0, 1\}^k$ and a vector $\mathbf{e} \in \{0, 1\}^r$. We make no assumptions about \mathbf{u} at this point, but we assume that \mathbf{e} is selected by Alice uniformly at random from $\{0, 1\}^r$. The encoder first constructs the vector $\mathbf{v} \in \{0, 1\}^n$, by setting $\mathbf{v}_{\mathcal{R}} = \mathbf{e}$, $\mathbf{v}_{\mathcal{A}} = \mathbf{u}$, and $\mathbf{v}_{\mathcal{B}} = \mathbf{0}$. The encoder then outputs $\mathcal{E}(\mathbf{u}, \mathbf{e}) := \mathbf{v} G_n = \mathbf{v} P_n G^{\otimes m}$ as in (5).

Decoding Algorithm: Formally, the decoder is a function $\mathcal{D}: \mathcal{Y}^n \rightarrow \{0, 1\}^k$. It accepts as input a vector $\mathbf{y} \in \mathcal{Y}^n$ at the output of the main channel $C_1 = \langle \{0, 1\}, \mathcal{Y}, W^* \rangle$. It then invokes successive cancellation decoding for the polar code $C_n(\mathcal{A} \cup \mathcal{R})$, used over W^* , to produce the vector $\hat{\mathbf{v}} \in \{0, 1\}^n$. The decoder outputs $\mathcal{D}(\mathbf{y}) := \hat{\mathbf{v}}_{\mathcal{A}}$.

We defer the proof that this coding scheme satisfies the reliability and security conditions (1), (2) to the next section. The rest of this section is devoted to two remarks about our construction.

Remark. We point out that our encoding algorithm can be regarded as a special case of the coset-coding method described in Section I-A. Recall that the coset-coding scheme is based upon an outer code \mathbf{C}^* that provides error-correction on the main channel and an inner code $\mathbf{C} \subset \mathbf{C}^*$ that ensures security for the wiretap channel. In our encoding algorithm, the outer code is $\mathbf{C}^* = C_n(\mathcal{A} \cup \mathcal{R})$ and the inner code is $\mathbf{C} = C_n(\mathcal{R})$. \square

Remark. Given the channel polarization phenomenon, it's intuitively clear from (25) why the proposed coding scheme should work. The information bits $\mathbf{U} = (U_1, U_2, \dots, U_k)$ reach Bob via good (almost noiseless) bit-channels. Thus Bob should be able to reconstruct them with very high probability. On the other hand, these same bits pass through bad (almost pure-noise) bit-channels on their way to Eve. Thus Eve should not be able to deduce much information about \mathbf{U} from her observations \mathbf{Z} , and $H(\mathbf{U}|\mathbf{Z})$ should be close to $H(\mathbf{U})$.

However, this simple intuition is misleading, because it does not show how the random bits in (25) help keep Eve ignorant. It may appear that this randomness is not really needed. For example, what would happen if the vector \mathbf{e} that serves as the second input to our encoder function $\mathcal{E}(\cdot, \cdot)$ is not chosen at random from $\{0, 1\}^r$ but rather set to an a priori fixed value? Since the channels are symmetric, any fixed value is as good as any other, so we may as well assume $\mathbf{e} = \mathbf{0}$. This does not seem to affect the argument in the foregoing paragraph and, according to this argument, $H(\mathbf{U}|\mathbf{Z})$ would still be close to $H(\mathbf{U})$.

In fact, this is not true. The reason is that channels seen by individual input bits as they undergo the transformation in (5) *depend on the distribution of other input bits*. Specifically, if $\mathbf{e} = \mathbf{0}$ or \mathbf{e} is fixed, the resulting encoder will *not* be secure. This is an important point that we would like to establish rigorously. To do so, we first need the following simple lemma.

Lemma 5. Let \mathcal{S} be an arbitrary subset of $[n]$ of size k , and suppose that the polar code $C_n(\mathcal{S})$ is used to communicate over a BSM channel $\langle \{0, 1\}, \mathcal{Z}, W \rangle$. Further, assume that the message \mathbf{U} at the input to the encoder for $C_n(\mathcal{S})$ is uniformly random over $\{0, 1\}^k$, and let \mathbf{Z} denote the random vector at the channel output. Then $I(\mathbf{U}; \mathbf{Z}) \geq k\mathcal{C}(W)$.

Proof. In fact, the lemma is true not only for $C_n(\mathcal{S})$ but for any binary linear code of dimension k . The only property of the transform matrix G_n that we need is that it is nonsingular.

Let \mathbf{V} be the random vector obtained by setting $\mathbf{V}_{\mathcal{S}} = \mathbf{U}$ and $\mathbf{V}_{\mathcal{S}^c} = \mathbf{0}$. Then the codeword transmitted over the channel is

$$\mathbf{X} = \mathbf{V} G_n = \mathbf{U} M \quad (29)$$

where M is a $k \times n$ row submatrix of G_n . Since G_n is nonsingular, $\text{rank}(M) = k$ and there exists a subset \mathcal{T} of $[n]$ of size k such that the corresponding k columns of M are linearly independent. This implies that there is a one-to-one correspondence between \mathbf{U} and $\mathbf{X}_{\mathcal{T}}$. Hence, $I(\mathbf{U}; \mathbf{Z}) = I(\mathbf{X}_{\mathcal{T}}; \mathbf{Z})$. Furthermore, since the random vector \mathbf{U} is uniform over $\{0, 1\}^k$, so is the vector $\mathbf{X}_{\mathcal{T}}$. Equivalently, its components $\{X_i : i \in \mathcal{T}\}$ are i.i.d. $\text{Ber}(1/2)$ random variables, and we can further conclude that

$$I(\mathbf{X}_{\mathcal{T}}; \mathbf{Z}) \geq I(\mathbf{X}_{\mathcal{T}}; \mathbf{Z}_{\mathcal{T}}) = \sum_{i \in \mathcal{T}} I(X_i; Z_i) = k\mathcal{C}(W)$$

where the last two equalities follow from the fact that the channel $\langle \{0, 1\}, \mathcal{Z}, W \rangle$ is memoryless and symmetric. \blacksquare

Now suppose that the input to our encoder function $\mathcal{E}(\cdot, \cdot)$ is a message chosen uniformly at random from $\{0, 1\}^k$ along with $\mathbf{e} = \mathbf{0}$. This is a special case of the situation considered in Lemma 5, with the set \mathcal{S} given by (27). Hence $I(\mathbf{U}; \mathbf{Z}) \geq k\mathcal{C}(W)$, and security condition (2) cannot be satisfied: a significant fraction of message bits, at least $\mathcal{C}(W)$, is potentially exposed. \square

We note that the foregoing remark illustrates a *general result*. It is known that (2) cannot be satisfied unless the encoder makes use of at least $I(\mathbf{X}; \mathbf{Z})$ random bits, where $I(\mathbf{X}; \mathbf{Z})$ is the mutual information between the input and output of Eve's channel [27].

V. WEAK SECURITY

In this section, we prove that the coding scheme of the previous section satisfies the reliability and security conditions (1) and (2) while its rate k/n approaches the secrecy capacity.

The reliability of this coding scheme follows immediately from Proposition 3. Let $\hat{\mathbf{V}}$ denote the random vector at the output of the successive cancellation decoder for $\mathbf{C}_n(\mathcal{A} \cup \mathcal{R})$ invoked by our decoding algorithm, and note that the corresponding probability of block error is upper bounded by (16). Since $\hat{\mathbf{U}} = \hat{\mathbf{V}}_{\mathcal{A}}$ and $\mathcal{A} \cup \mathcal{R} = \mathcal{G}_n(W^*, \beta)$ by design, we see that

$$\Pr\{\hat{\mathbf{U}} \neq \mathbf{U}\} \leq \sum_{i \in \mathcal{A} \cup \mathcal{R}} Z(W_i^*) \leq 2^{-n^\beta} \quad (30)$$

Since $n \geq k$, this clearly implies that $\lim_{k \rightarrow \infty} \Pr\{\hat{\mathbf{U}} \neq \mathbf{U}\} = 0$ as required in (1), and the reliability condition is satisfied.

We now turn to the proof of security. For the remainder of this section, the sets \mathcal{R} , \mathcal{A} , \mathcal{B} are given by (26)–(28), \mathbf{U} denotes Alice's message, \mathbf{V} denotes the intermediate vector constructed by our encoding algorithm (with $\mathbf{V}_{\mathcal{A}} = \mathbf{U}$, $\mathbf{V}_{\mathcal{B}} = \mathbf{0}$, and $\mathbf{V}_{\mathcal{R}}$ uniform over $\{0, 1\}^r$), and \mathbf{Z} denotes Eve's observations. Also $|\mathcal{A}| = k$, $|\mathcal{R}| = r$, and $h_2(\cdot)$ is the binary entropy function.

Lemma 6.

$$H(\mathbf{V}_{\mathcal{R}}|\mathbf{Z}, \mathbf{V}_{\mathcal{A}}) \leq h_2(2^{-n^\beta}) + r 2^{-n^\beta}$$

Proof. Suppose that in addition to her observations \mathbf{Z} , a genie reveals to Eve the realization $\mathbf{v}_{\mathcal{A}}$ of $\mathbf{V}_{\mathcal{A}}$ and asks her to produce an estimate of $\mathbf{V}_{\mathcal{R}}$. Eve also knows that $\mathbf{V}_{\mathcal{B}} = \mathbf{0}$. Thus she knows all the bits of $\mathbf{V}_{\mathcal{R}^c}$. Since $\mathcal{R} = \mathcal{G}_n(W, \beta)$, this is precisely the scenario considered in (24). Consequently, Eve can use successive cancellation decoding to deterministically compute an estimate $\hat{\mathbf{V}}_{\mathcal{R}} = f(\mathbf{Z}, \mathbf{V}_{\mathcal{A}})$ such that

$$\lambda \stackrel{\text{def}}{=} \Pr\{\hat{\mathbf{V}}_{\mathcal{R}} \neq \mathbf{V}_{\mathcal{R}}\} \leq \sum_{i \in \mathcal{R}} Z(W_i) \leq 2^{-n^\beta} \quad (31)$$

We now invoke Fano's inequality [11, p. 38] to bound the conditional entropy $H(\mathbf{V}_{\mathcal{R}}|\mathbf{Z}, \mathbf{V}_{\mathcal{A}})$ in terms of λ as follows

$$H(\mathbf{V}_{\mathcal{R}}|\mathbf{Z}, \mathbf{V}_{\mathcal{A}}) \leq h_2(\lambda) + r\lambda \quad (32)$$

where we have also used the fact that $\mathbf{V}_{\mathcal{R}}$ takes values in the set $\{0, 1\}^r$ of size 2^r . The lemma now follows from (31), (32), and the fact that $h_2(\cdot)$ is increasing on the interval $[0, 1/2]$. ■

Let us define $\epsilon_n \stackrel{\text{def}}{=} \mathcal{C}(W) - |\mathcal{R}|/n$. Since $\mathcal{R} = \mathcal{G}_n(W, \beta)$, Theorem 1 implies that $\lim_{n \rightarrow \infty} \epsilon_n = 0$.

Lemma 7.

$$I(\mathbf{U}; \mathbf{Z}) \leq n\epsilon_n + h_2(2^{-n^\beta}) + (n - k) 2^{-n^\beta} \quad (33)$$

Proof. The lemma is proved via a long sequence of simple equalities and inequalities, as follows:

$$I(\mathbf{U}; \mathbf{Z}) = I(\mathbf{V}_{\mathcal{A}}; \mathbf{Z}) = I(\mathbf{V}_{\mathcal{A} \cup \mathcal{B}}; \mathbf{Z}) \quad (34)$$

$$= I(\mathbf{V}; \mathbf{Z}) - I(\mathbf{V}_{\mathcal{R}}; \mathbf{Z}|\mathbf{V}_{\mathcal{A} \cup \mathcal{B}}) \quad (35)$$

$$= I(\mathbf{V}; \mathbf{Z}) - I(\mathbf{V}_{\mathcal{R}}; \mathbf{Z}|\mathbf{V}_{\mathcal{A}}) \quad (36)$$

$$= I(\mathbf{V}; \mathbf{Z}) - H(\mathbf{V}_{\mathcal{R}}|\mathbf{V}_{\mathcal{A}}) + H(\mathbf{V}_{\mathcal{R}}|\mathbf{Z}, \mathbf{V}_{\mathcal{A}}) \quad (37)$$

$$= I(\mathbf{V}; \mathbf{Z}) - H(\mathbf{V}_{\mathcal{R}}) + H(\mathbf{V}_{\mathcal{R}}|\mathbf{Z}, \mathbf{V}_{\mathcal{A}}) \quad (38)$$

$$= I(\mathbf{V}; \mathbf{Z}) - r + H(\mathbf{V}_{\mathcal{R}}|\mathbf{Z}, \mathbf{V}_{\mathcal{A}}) \quad (39)$$

$$\leq n\mathcal{C}(W) - r + H(\mathbf{V}_{\mathcal{R}}|\mathbf{Z}, \mathbf{V}_{\mathcal{A}}) \quad (40)$$

$$= n\epsilon_n + H(\mathbf{V}_{\mathcal{R}}|\mathbf{Z}, \mathbf{V}_{\mathcal{A}}) \quad (41)$$

$$\leq n\epsilon_n + h_2(2^{-n^\beta}) + r 2^{-n^\beta} \quad (42)$$

$$\leq n\epsilon_n + h_2(2^{-n^\beta}) + (n - k) 2^{-n^\beta} \quad (43)$$

The equalities in (34) hold since $\mathbf{V}_{\mathcal{A}} = \mathbf{U}$ and $\mathbf{V}_{\mathcal{B}} = \mathbf{0}$. Now observe that $\mathcal{A} \cup \mathcal{B}$ and \mathcal{R} are complements of each other in $[n]$. Hence, any distribution of \mathbf{V} can be thought of as a joint distribution of $\mathbf{V}_{\mathcal{R}}$ and $\mathbf{V}_{\mathcal{A} \cup \mathcal{B}}$. Given this observation, (35) follows from the chain rule for mutual information. The equality in (36) is trivial from $\mathbf{V}_{\mathcal{B}} = \mathbf{0}$, while (37) is the definition of conditional mutual information. The equalities (38) and (39) hold since $\mathbf{V}_{\mathcal{R}}$ and $\mathbf{V}_{\mathcal{A}}$ are independent, and $\mathbf{V}_{\mathcal{R}}$ is a priori uniform over $\{0, 1\}^r$. Inequality (40) is immediate from the fact that $\mathcal{C}(W)$ is the capacity of W , while (41) follows from the definition of ϵ_n . Finally, (42) follows from Lemma 6 and (43) is trivial. ■

Theorem 8. *The encoding algorithm of the previous section satisfies the weak security condition (2), namely*

$$\lim_{k \rightarrow \infty} \frac{I(\mathbf{U}; \mathbf{Z})}{k} = 0$$

Proof. This follows immediately from Lemma 7. Divide both sides of (33) by n to get

$$\frac{I(\mathbf{U}; \mathbf{Z})}{n} \leq \epsilon_n + \frac{h_2(2^{-n^\beta})}{n} + \frac{n - k}{n} 2^{-n^\beta} \quad (44)$$

It is clear that the last two terms in (44) tend to zero as $n \rightarrow \infty$, and $\lim_{n \rightarrow \infty} \epsilon_n = 0$ by Theorem 1. Along with the obvious fact that $k = \Omega(n)$, this completes the proof of the theorem. ■

Recall that for the wiretap-channel systems considered in this paper, the secrecy capacity is $\mathcal{C}_s = \mathcal{C}(W^*) - \mathcal{C}(W)$ (cf. Section II). Let $R_n = k/n$ denote the rate of our coding scheme.

Theorem 9.

$$\lim_{n \rightarrow \infty} R_n = \mathcal{C}(W^*) - \mathcal{C}(W)$$

Proof. Observe that

$$R_n = \frac{|\mathcal{A}|}{n} = \frac{|\mathcal{G}_n(W^*, \beta)|}{n} - \frac{|\mathcal{G}_n(W, \beta)|}{n} \quad (45)$$

where we have used the definition of \mathcal{A} in (27) and the fact that $\mathcal{G}_n(W, \beta)$ is a subset of $\mathcal{G}_n(W^*, \beta)$ for all $\beta < 1/2$. The theorem now follows from Theorem 1. ■

Theorem 9 does not directly imply that our coding scheme achieves secrecy capacity, because the rate of communication from Alice to Bob, measured in information bits per channel use, could be much less than k/n when $H(\mathbf{U}) < k$. But this is true for *any* encoder that converts k input bits to n coded output bits. If the encoder accommodates an arbitrary distribution on its input \mathbf{U} , it can achieve capacity *only* when $H(\mathbf{U}) = k(1 - o(1))$. If this necessary condition is satisfied, then our coding scheme does achieve secrecy capacity by Theorem 9.

VI. STRONG SECURITY

This section shows how polar coding could be used to provide strong security whenever the main channel C_1 and the wiretap channel C_2 are symmetric binary-input DMCs, and C_2 is degraded with respect to C_1 . Specifically, we describe a polar coding scheme that satisfies the strong security condition (3), while operating at a rate k/n that approaches the secrecy capacity.

First, we will introduce a subtle but important change in the coding scheme of Section IV (see Section VI-B). In order to show that this change suffices to guarantee strong security, we need to replace the proof in the previous section by a more intricate argument (see Section VI-D). This argument relies crucially on the fact that a certain composite channel induced by our construction is symmetric (Section VI-C). Aided by a recent result of Hassani and Urbanke [16], we then prove that the rate of the proposed coding scheme approaches the secrecy capacity (Section VI-E). Unfortunately, we can show that the reliability condition (1) is satisfied only for the case where the main channel is noiseless. Nevertheless, we believe that, in practice, low probabilities of block error can be achieved also when the main channel is not noiseless, using a modification of Arkan's successive cancellation decoder (Section VI-F).

A. Analysis of the Weak-Security Coding Scheme

Henceforth, let us refer to the coding scheme introduced in Section IV and analyzed in the previous section as the **weak-security coding scheme**. A natural question is whether this coding scheme does, in fact, provide strong security. Although we do not have a definitive answer to this question, we conjecture that it does not. As before, let

$$\epsilon_n \stackrel{\text{def}}{=} \mathcal{C}(W) - \frac{|\mathcal{R}|}{n} = \mathcal{C}(W) - \frac{|\mathcal{G}_n(W, \beta)|}{n} \quad (46)$$

with the sets \mathcal{R} and $\mathcal{G}_n(W, \beta)$ defined as in (26) and (13). The following result provides some evidence for this conjecture.

Proposition 10. *Whenever the wiretap channel is a binary-input symmetric DMC and the main channel is noiseless, the weak-security coding scheme achieves strong security if and only if*

$$\lim_{n \rightarrow \infty} n\epsilon_n = 0 \quad (47)$$

Proof. The fact that (47) is sufficient for strong security is obvious from Lemma 7, since the last two terms in (33) tend to zero exponentially fast. In fact, it is clear that (47) is sufficient for strong security, whether the main channel is noiseless or not. We now show that this condition is also necessary, at least in the case where the main channel is noiseless. In this case, the set \mathcal{B} in (28) is empty, and the vector \mathbf{V} at the input to the transformation (5) consists of $\mathbf{V}_A = \mathbf{U}$ and \mathbf{V}_R , with \mathbf{V}_R being uniform over $\{0, 1\}^r$. Hence, if the message \mathbf{U} is a priori uniform over $\{0, 1\}^k$, then \mathbf{V} is uniform over $\{0, 1\}^n$. Let $\mathbf{X} = \mathbf{V}\mathbf{G}_n$ as in (5). Since the Arkan transform matrix \mathbf{G}_n is nonsingular, \mathbf{X} is also uniform over $\{0, 1\}^n$. Consequently, we have

$$I(\mathbf{V}; \mathbf{Z}) = I(\mathbf{X}; \mathbf{Z}) = \sum_{i=1}^n I(X_i, Z_i) = n\mathcal{C}(W) \quad (48)$$

where the second equality follows by noting that X_1, X_2, \dots, X_n are i.i.d. $\text{Ber}(1/2)$ random variables and W is memoryless, while

the last equality follows from the fact that W is symmetric. This implies that the inequality (40) in the proof of Lemma 7 becomes an equality in this case. Therefore

$$I(\mathbf{U}; \mathbf{Z}) = n\epsilon_n + H(\mathbf{V}_R | \mathbf{Z}, \mathbf{V}_A) \geq n\epsilon_n \quad (49)$$

It is now clear that $\lim_{n \rightarrow \infty} n\epsilon_n = 0$ is necessary for the mutual information $I(\mathbf{U}; \mathbf{Z})$ to vanish asymptotically. ■

Given a BSM channel W , is it true that $\lim_{n \rightarrow \infty} n\epsilon_n = 0$ for this channel? Unfortunately, the answer to this question is negative. It is known [34, 36] that for any discrete memoryless channel W and any code of length n and rate R that achieves error probability P_e on W , we have

$$\mathcal{C}(W) - R \geq \frac{\text{const}(P_e, W)}{\sqrt{n}} - O\left(\frac{\log n}{n}\right)$$

where the constant (which is given explicitly in [34]) depends on W and P_e , but not on n . This implies that $n\epsilon_n = \Omega(\sqrt{n})$, and the weak-security coding scheme does *not* provide strong security. Consequently, in order to provide strong security, the polar coding scheme of Section IV has to be modified.

B. Strong-Security Coding Scheme

Intuitively, the main reason that the coding scheme of Section IV fails to provide strong security is this: the bit-channels that are deemed bad for Eve are not bad enough. Indeed, according to the definition of $\mathcal{B}_n(W, \beta)$ in (14), a bit-channel W_i is considered bad for Eve whenever $Z(W_i) \geq 2^{-n^\beta}/n$. For example, if $n = 2^{10}$ and $\beta = 0.499$, a bit-channel may be declared bad for Eve even when its capacity is greater than $1 - 10^{-9}$ while Eve's probability of error on this channel is less than $2 \cdot 10^{-10}$. It is obvious that such a bit-channel does not prevent Eve from deducing the information at its input with high probability.

The problem is that the generally accepted definitions of good and bad bit-channels — for example, (13) and (14) — are motivated by *Bob's point of view*. First, a criterion for “goodness” is established, motivated by the probability of error on Bob's side, then the bit-channels that do not satisfy this criterion are deemed bad. In order to achieve strong security, we will re-define things from *Eve's point of view*. First, we introduce a strong criterion for “badness,” and then make sure that random bits are sent over all the bit-channels that do not satisfy this criterion. Specifically, given a BSM channel W and a positive $\delta < 1$, we define the index set of δ -**poor bit-channels** as follows:

$$\mathcal{P}_n(W, \delta) \stackrel{\text{def}}{=} \{i \in [n] : \mathcal{C}(W_i) \leq \delta\} \quad (50)$$

Further, we leave the definition of the good bit-channels in (13) unchanged, but re-define the sets \mathcal{R} , \mathcal{A} , and \mathcal{B} as follows:

$$\mathcal{R} \stackrel{\text{def}}{=} [n] \setminus \mathcal{P}_n(W, \delta_n) \quad (51)$$

$$\mathcal{A} \stackrel{\text{def}}{=} \mathcal{P}_n(W, \delta_n) \cap \mathcal{G}_n(W^*, \beta) \quad (52)$$

$$\mathcal{B} \stackrel{\text{def}}{=} \mathcal{P}_n(W, \delta_n) \setminus \mathcal{G}_n(W^*, \beta) \quad (53)$$

where, for the time being, δ_n is an arbitrary function from the positive integers to the interval $(0, 1)$. We will specify this function precisely later in this section (cf. Theorem 17 and Proposi-

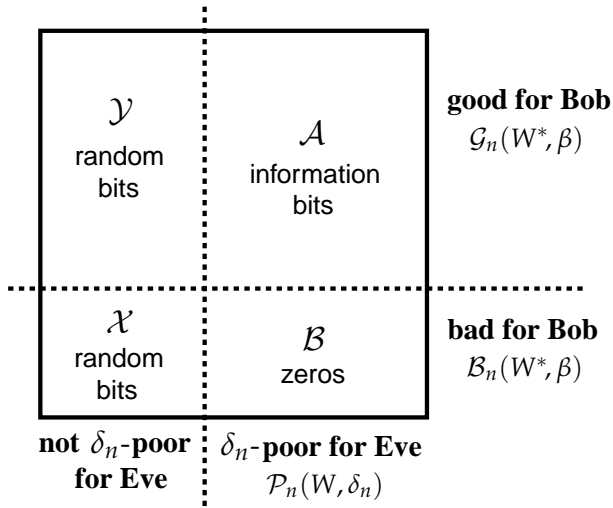


Figure 2. Informal sketch of the strong-security coding scheme

tion 20). In the meantime, notice that the sets \mathcal{R} , \mathcal{A} , and \mathcal{B} , as defined in (51)–(53), are still disjoint and $\mathcal{R} \cup \mathcal{A} \cup \mathcal{B} = [n]$.

The **strong-security coding scheme** of (51)–(53) is illustrated schematically in Figure 2, wherein the bold square represents the set of n bit-channels. This set is partitioned two ways: into bit-channels that are good for Bob and bad for Bob (as before), and into bit-channels that are δ_n -poor and not δ_n -poor for Eve. The sets \mathcal{X} and \mathcal{Y} , with $\mathcal{X} \cup \mathcal{Y} = \mathcal{R}$, will be discussed later in this section (see (94), (95) and Section VI-F).

With the new definitions of the sets \mathcal{R} , \mathcal{A} , \mathcal{B} in (51)–(53), the encoding and decoding algorithms for the strong-security coding scheme are exactly those given in Section IV for the weak-security coding scheme (although we will modify the decoding algorithm somewhat in Section VI-F).

C. The Induced Channel Is Symmetric

We prove in the next subsection that the strong-security coding scheme indeed provides strong security. In order to do so, we first introduce and study a certain composite channel induced by our construction. Informally, this channel describes the transformation in (5) in the case where some r of the bits V_1, V_2, \dots, V_n are set independently and uniformly at random, while the remaining $n - r$ bits serve as the input to the channel. This situation is depicted in Figure 3. Formally, the **induced channel** $\mathcal{Q}_n(W, \mathcal{R})$ is specified in terms of an arbitrary BSM channel W with output alphabet \mathcal{Z} , and a subset \mathcal{R} of $[n]$ of size $|\mathcal{R}| = r$. The input alphabet of $\mathcal{Q}_n(W, \mathcal{R})$ is $\{0, 1\}^{n-r}$ and its output alphabet is \mathcal{Z}^n . To describe the transition probabilities of $\mathcal{Q}_n(W, \mathcal{R})$, let us introduce the following notation. Henceforth, given a vector $\mathbf{x} \in \{0, 1\}^{n-r}$ and a vector $\mathbf{e} \in \{0, 1\}^r$, let $(\mathbf{x}; \mathbf{e})$ denote the vector $\mathbf{v} \in \{0, 1\}^n$ with $\mathbf{v}_{\mathcal{R}} = \mathbf{e}$ and $\mathbf{v}_{\mathcal{R}^c} = \mathbf{x}$. With this, referring to the definition of W^n in (9), the $2^{n-r} \times |\mathcal{Z}|^n$ transition-probability matrix Q of $\mathcal{Q}_n(W, \mathcal{R})$ is given by

$$Q(\mathbf{z}|\mathbf{x}) \stackrel{\text{def}}{=} \frac{1}{2^r} \sum_{\mathbf{e} \in \{0,1\}^r} W^n(\mathbf{z} | (\mathbf{x}; \mathbf{e}) G_n) \quad (54)$$

for all $\mathbf{x} \in \{0, 1\}^{n-r}$ and all $\mathbf{z} \in \mathcal{Z}^n$. It can be readily seen that if $\mathbf{V} = (V_1, V_2, \dots, V_{n-r})$ is the random vector at the input to

the channel in Figure 3 and $\mathbf{Z} = (Z_1, Z_2, \dots, Z_n)$ is the random vector at the channel output, then indeed

$$Q(\mathbf{z}|\mathbf{x}) = \Pr\{\mathbf{Z} = \mathbf{z} | \mathbf{V} = \mathbf{x}\} \quad (55)$$

Our main goal in this subsection is to show that the induced channel $\mathcal{Q}_n(W, \mathcal{R})$ in Figure 3 is symmetric. This follows as a special case of the more general results in [5]. Although [5] precedes this paper chronologically, it is not yet publicly available. Therefore, we include a complete proof for completeness.

Recall that a **group action** of an abelian group A on a set \mathcal{Y} is a function from $A \times \mathcal{Y}$ to \mathcal{Y} , denoted $(a, \mathbf{y}) \mapsto a \cdot \mathbf{y}$, with the following properties:

- P1. $0 \cdot \mathbf{y} = \mathbf{y}$ for all $\mathbf{y} \in \mathcal{Y}$, where 0 is the identity of A ;
- P2. $(a + b) \cdot \mathbf{y} = a \cdot (b \cdot \mathbf{y})$ for all $a, b \in A$ and all $\mathbf{y} \in \mathcal{Y}$, where $+$ denotes the group operation.

The orbit of $\mathbf{y} \in \mathcal{Y}$ is the set of all points of \mathcal{Y} to which \mathbf{y} can be moved by the elements of A . Explicitly, the **orbit of \mathbf{y}** is

$$\mathcal{O}(\mathbf{y}) \stackrel{\text{def}}{=} \{a \cdot \mathbf{y} : a \in A\} \quad (56)$$

It is well known that orbits of points in \mathcal{Y} form a partition of \mathcal{Y} into equivalence classes (under the equivalence relation $\mathbf{y}_1 \sim \mathbf{y}_2$ iff there exists an $a \in A$ with $\mathbf{y}_2 = a \cdot \mathbf{y}_1$).

Theorem 11. Let $\langle \mathcal{X}, \mathcal{Y}, W \rangle$ be a DMC, and suppose that \mathcal{X} is an abelian group under the binary operation $+$. Further, suppose that there exists a group action \cdot of \mathcal{X} on \mathcal{Y} such that

$$W(\mathbf{y}|a + x) = W(a \cdot \mathbf{y}|x) \quad (57)$$

for all $a, x \in \mathcal{X}$ and all $\mathbf{y} \in \mathcal{Y}$. Then the DMC $\langle \mathcal{X}, \mathcal{Y}, W \rangle$ is necessarily a symmetric channel.

Proof. We partition the set \mathcal{Y} into orbits formed by the group action \cdot of \mathcal{X} . Let \mathcal{O} be an orbit, and let M be the $|\mathcal{X}| \times |\mathcal{O}|$ column submatrix of W consisting of those columns that are indexed by the elements of \mathcal{O} . It would suffice to prove that M is a strongly symmetric matrix, regardless of the choice of \mathcal{O} .

Consider two arbitrary rows of M indexed, say, by the elements $x_1 \in \mathcal{X}$ and $x_2 \in \mathcal{X}$. Set $a = x_2 + (-x_1)$, where $-x_1$ is the inverse of x_1 in the group \mathcal{X} . Then we have $x_2 = a + x_1$. Therefore, (57) implies that

$$W(\mathbf{y}|x_2) = W(a \cdot \mathbf{y}|x_1) \quad \text{for all } \mathbf{y} \in \mathcal{Y} \quad (58)$$

It is easy to see from (56) and property P2 that the map $\mathbf{y} \mapsto a \cdot \mathbf{y}$ is bijective on \mathcal{O} . Together with (58), this shows that the rows of M , indexed by x_1 and x_2 , are permutations of each other.

Now consider two arbitrary columns of M indexed by the elements $\mathbf{y}_1 \in \mathcal{O}$ and $\mathbf{y}_2 \in \mathcal{O}$. Then, by the definition of an orbit, there exists an $a \in \mathcal{X}$ such that $\mathbf{y}_2 = a \cdot \mathbf{y}_1$, and (57) implies that

$$W(\mathbf{y}_2|x) = W(\mathbf{y}_1|a + x) \quad \text{for all } x \in \mathcal{X} \quad (59)$$

It is clear that the map $x \mapsto a + x$ is bijective on \mathcal{X} . Thus (59) shows that the columns of M are permutations of each other. ■

Notice that the input alphabet $\{0, 1\}^{n-r}$ of $\mathcal{Q}_n(W, \mathcal{R})$ is an abelian group, with the group operation $+$ being the componentwise modulo 2 addition of vectors in \mathbb{F}_2^{n-r} . Consequently, in order to prove that the channel $\mathcal{Q}_n(W, \mathcal{R})$ is symmetric, it would suffice to construct a group action of $\{0, 1\}^{n-r}$ on \mathcal{Z}^n

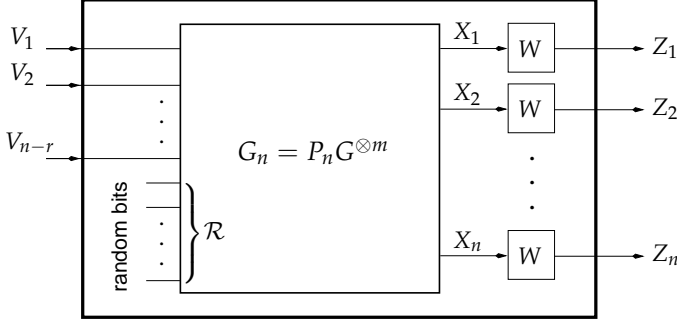


Figure 3. Block diagram of the induced channel $\mathcal{Q}_n(W, \mathcal{R})$

that satisfies (57). To do so, we will use the fact that W itself is symmetric. As noted by Arıkan [3], a binary-input channel W with output alphabet \mathcal{Z} is symmetric if and only if there exists a permutation π_1 on \mathcal{Z} such that

$$\pi_1 = \pi_1^{-1} \quad (\pi_1 \text{ is an involution}) \quad (60)$$

$$W(z|0) = W(\pi_1(z)|1) \quad \text{for all } z \in \mathcal{Z} \quad (61)$$

Let π_0 be the identity permutation on \mathcal{Z} . Following Arıkan [3], let us define a group action of the additive group of $\mathbb{F}_2 = \{0, 1\}$ on the set \mathcal{Z} as follows: $x.z = \pi_x(z)$ for all $x \in \mathbb{F}_2$ and $z \in \mathcal{Z}$. It is trivial to verify that $x.z$ has the required group-action properties P1 and P2, and that for all $a, x \in \mathbb{F}_2$ and $z \in \mathcal{Z}$, we have

$$W(z|a+x) = W(a.z|x) \quad (62)$$

As in [3], we can extend this function componentwise to a group action of the additive group of \mathbb{F}_2^n on the set \mathcal{Z}^n as follows:

$$\mathbf{x}.z \stackrel{\text{def}}{=} (x_1.z_1, x_2.z_2, \dots, x_n.z_n) \quad (63)$$

The following lemma was proved by Arıkan in [3, Propositions 12 and 13]. We provide a simple proof herein, for completeness.

Lemma 12. Let $G_n = P_n G^{\otimes m}$ be the Arıkan transform matrix. Then for all $\mathbf{a}, \mathbf{x} \in \mathbb{F}_2^n$ and all $\mathbf{z} \in \mathcal{Z}^n$, we have

$$W^n(\mathbf{z} | (\mathbf{a} + \mathbf{x}) G_n) = W^n(\mathbf{a} G_n . \mathbf{z} | \mathbf{x} G_n) \quad (64)$$

Proof. In fact, the lemma is true for an arbitrary $n \times n$ binary matrix (in other words, any linear transformation from \mathbb{F}_2^n to itself). First, let us show that

$$W^n(\mathbf{z} | \mathbf{b} + \mathbf{c}) = W^n(\mathbf{b}.z | \mathbf{c}) \quad (65)$$

for all $\mathbf{b}, \mathbf{c} \in \mathbb{F}_2^n$ and all $\mathbf{z} \in \mathcal{Z}^n$. This follows directly from the definition of W^n . Indeed, expanding both sides of (65) as

$$\prod_{i=1}^n W(z_i | b_i + c_i) = \prod_{i=1}^n W(b_i.z_i | c_i)$$

we conclude that (65) is implied by (62). Let us now set $\mathbf{b} = \mathbf{a} G_n$ and $\mathbf{c} = \mathbf{x} G_n$. Then the lemma follows from (65) along with the fact that multiplication by a matrix is a linear operation, that is $(\mathbf{a} + \mathbf{x}) G_n = \mathbf{a} G_n + \mathbf{x} G_n = \mathbf{b} + \mathbf{c}$. ■

We now depart from Arıkan [3], and introduce a group action \circ of the additive group of \mathbb{F}_2^{n-r} on \mathcal{Z}^n , defined as follows. Recall that (\mathbf{x}, \mathbf{e}) denotes the vector $\mathbf{v} \in \{0, 1\}^n$ with $\mathbf{v}_R = \mathbf{e}$ and $\mathbf{v}_{R^c} = \mathbf{x}$. With this, we define for all $\mathbf{x} \in \mathbb{F}_2^{n-r}$ and all $\mathbf{z} \in \mathcal{Z}^n$

$$\mathbf{x} \circ \mathbf{z} \stackrel{\text{def}}{=} (\mathbf{x}; \mathbf{0}) G_n . \mathbf{z} \quad (66)$$

where the group action $.$ on the right-hand side is the one defined in (63). Again, it is easy to verify that (66) satisfies P1 and P2. Our main result in this subsection is the following proposition.

Proposition 13. The induced channel $\mathcal{Q}_n(W, \mathcal{R})$ is symmetric.

Proof. In light of Theorem 11, it would suffice to prove that the transition-probability matrix Q defined in (54) satisfies

$$Q(\mathbf{z} | \mathbf{a} + \mathbf{x}) = Q(\mathbf{a} \circ \mathbf{z} | \mathbf{x})$$

for all $\mathbf{a}, \mathbf{x} \in \mathbb{F}_2^{n-r}$ and $\mathbf{z} \in \mathcal{Z}^n$. Expanding the vector $(\mathbf{a} + \mathbf{x}; \mathbf{e})$ as $(\mathbf{a}; \mathbf{0}) + (\mathbf{x}; \mathbf{e})$, and substituting in (54), we obtain

$$Q(\mathbf{z} | \mathbf{a} + \mathbf{x}) = \frac{1}{2^r} \sum_{\mathbf{e} \in \{0,1\}^r} W^n(\mathbf{z} | ((\mathbf{a}; \mathbf{0}) + (\mathbf{x}; \mathbf{e})) G_n) \quad (67)$$

$$= \frac{1}{2^r} \sum_{\mathbf{e} \in \{0,1\}^r} W^n((\mathbf{a}; \mathbf{0}) G_n . \mathbf{z} | (\mathbf{x}; \mathbf{e}) G_n) \quad (68)$$

$$= \frac{1}{2^r} \sum_{\mathbf{e} \in \{0,1\}^r} W^n(\mathbf{a} \circ \mathbf{z} | (\mathbf{x}; \mathbf{e}) G_n) \quad (69)$$

where (68) follows from Lemma 12 and (69) follows from (66). But (69) is precisely $Q(\mathbf{a} \circ \mathbf{z} | \mathbf{x})$ by (54), and we are done. ■

D. Proof of Strong Security

Let us begin with a simple lemma that relates the capacity of the bit-channels in (11) to the transformation in (5). Although this lemma is well-known, we include a proof for completeness.

Lemma 14. Let W be an arbitrary BSM channel. Suppose the vector $\mathbf{V} = (V_1, V_2, \dots, V_n)$ at the input to the transformation in (5) is uniform over $\{0, 1\}^n$, and let $\mathbf{Z} = (Z_1, Z_2, \dots, Z_n)$ be the random vector at the output of the transformation. Further, let W_1, W_2, \dots, W_n be the corresponding bit-channels, defined in (11). Then for all $i \in [n]$, the capacity of W_i is given by

$$\mathcal{C}(W_i) = I(V_i; \mathbf{Z}, V_1, V_2, \dots, V_{i-1})$$

Proof. Let \mathbf{V}_i denote the random vector (V_1, V_2, \dots, V_i) for all $i \in [n]$, as before. It is shown in [3] that if W is symmetric, then so is W_i for all $i \in [n]$. Hence, the capacity of W_i is the mutual information between its input and output when the input is uniform over $\{0, 1\}$. Thus it would suffice to show that

$$W_i(\mathbf{z}, \mathbf{v} | x) = \Pr\{\mathbf{Z} = \mathbf{z}, \mathbf{V}_{i-1} = \mathbf{v} | V_i = x\} \quad (70)$$

for all $\mathbf{z} \in \mathcal{Z}^n$, $\mathbf{v} \in \{0, 1\}^{i-1}$, and $x \in \{0, 1\}$. Since V_i is uniform we can re-write the right-hand side of (70) as follows:

$$\frac{\Pr\{\mathbf{Z} = \mathbf{z}, \mathbf{V}_{i-1} = \mathbf{v}, V_i = x\}}{\Pr\{V_i = x\}} = 2 \Pr\{\mathbf{Z} = \mathbf{z}, V_i = (\mathbf{v}, x)\}$$

Observe that the event $\{V_i = (\mathbf{v}, x)\}$ is the union of 2^{n-i} disjoint events $\{V = (\mathbf{v}, x, \bar{\mathbf{v}})\}$, as $\bar{\mathbf{v}}$ ranges over $\{0, 1\}^{n-i}$ (or $\bar{\mathbf{v}}$

is the empty string, if $i = n$). Since \mathbf{V} is uniform over $\{0, 1\}^n$, the probability of each such event is 2^{-n} . Consequently

$$\begin{aligned} 2 \Pr\{\mathbf{Z} = \mathbf{z}, \mathbf{V}_i = (v, x)\} &= \\ &= 2 \sum_{\bar{v} \in \{0,1\}^{n-i}} \Pr\{\mathbf{Z} = \mathbf{z}, \mathbf{V} = (v, x, \bar{v})\} \end{aligned} \quad (71)$$

$$= \frac{1}{2^{n-1}} \sum_{\bar{v} \in \{0,1\}^{n-i}} \Pr\{\mathbf{Z} = \mathbf{z} \mid \mathbf{V} = (v, x, \bar{v})\} \quad (72)$$

Since \mathbf{Z} and \mathbf{V} are, respectively, the output and the input to the transformation in (5), for all $\mathbf{w} \in \{0, 1\}^n$ we have

$$\Pr\{\mathbf{Z} = \mathbf{z} \mid \mathbf{V} = \mathbf{w}\} = W^n(\mathbf{z} \mid \mathbf{w} P_n G^{\otimes m}) = \tilde{W}(\mathbf{z} \mid \mathbf{w})$$

by the definition of the “combined” channel \tilde{W} in (10). Together with (72) and the definition of W_i in (11), this shows that the right-hand side of (70) is indeed equal to $W_i(\mathbf{z}, v \mid x)$. ■

The next lemma combines Proposition 13 with Lemma 14 to upper-bound the capacity of the induced channel $\mathcal{Q}_n(W, \mathcal{R})$.

Lemma 15. *Let W be an arbitrary BSM channel. For $n = 2^m$, let W_1, W_2, \dots, W_n denote the corresponding bit-channels. Then for all $\mathcal{R} \subset [n]$, the capacity of the induced channel $\mathcal{Q}_n(W, \mathcal{R})$, defined in (54), is upper-bounded as follows:*

$$\mathcal{C}(\mathcal{Q}_n(W, \mathcal{R})) \leq \sum_{i \in \mathcal{R}^c} \mathcal{C}(W_i) \quad (73)$$

Proof. Consider again the transformation in (5), with the input vector $\mathbf{V} = (V_1, V_2, \dots, V_n)$ being uniform over $\{0, 1\}^n$, and let $\mathbf{Z} = (Z_1, Z_2, \dots, Z_n)$ denote the output of the transformation, as in Lemma 14. Then

$$\mathcal{C}(\mathcal{Q}_n(W, \mathcal{R})) = I(\mathbf{V}_{\mathcal{R}^c}; \mathbf{Z}) \quad (74)$$

This is so because $\mathcal{Q}_n(W, \mathcal{R})$ is symmetric by Proposition 13, and the capacity of a symmetric DMC is given by the mutual information between its input and output, *under uniform input distribution* [14, Theorem 4.5.2]. Next, write $\mathcal{R}^c = [n] \setminus \mathcal{R}$ as

$$\mathcal{R}^c = \{i_1, i_2, \dots, i_{n-r}\}$$

where $r = |\mathcal{R}|$, and assume w.l.o.g. that $i_1 < i_2 < \dots < i_{n-r}$. The lemma can be now proved via a sequence of simple equalities and inequalities, as follows:

$$I(\mathbf{V}_{\mathcal{R}^c}; \mathbf{Z}) = I(V_{i_1}, V_{i_2}, \dots, V_{i_{n-r}}; \mathbf{Z}) \quad (75)$$

$$= \sum_{j=1}^{n-r} I(V_{i_j}; \mathbf{Z} \mid V_{i_1}, V_{i_2}, \dots, V_{i_{j-1}}) \quad (76)$$

$$= \sum_{j=1}^{n-r} I(V_{i_j}; \mathbf{Z}, V_{i_1}, V_{i_2}, \dots, V_{i_{j-1}}) \quad (77)$$

$$\leq \sum_{j=1}^{n-r} I(V_{i_j}; \mathbf{Z}, V_1, V_2, \dots, V_{i_{j-1}}) \quad (78)$$

The equality (76) is the chain rule for mutual information. The equality (77) follows from the fact that $I(X; Z \mid Y) = I(X; Z, Y)$ for all random variables X, Y, Z such that X and Y are independent. To establish (78), we adjoin to the set of random variables $\{V_{i_1}, V_{i_2}, \dots, V_{i_{j-1}}\}$ its complement in the set $\{V_1, V_2, \dots, V_{i_j}\}$.

Clearly, this cannot decrease the mutual information. Lastly, we observe that the summation in (78) is equal to the summation on the right-hand side of (73) by Lemma 14. ■

With Lemma 15 in hand, we are finally ready to establish our main result in this section.

Proposition 16. *Let \mathbf{U} be the message at the input to the encoder for the strong-security coding scheme. Then, regardless of the a priori distribution of \mathbf{U} , we have*

$$I(\mathbf{U}; \mathbf{Z}) \leq \delta_n |\mathcal{P}_n(W, \delta_n)| \quad (79)$$

where \mathbf{Z} is the output of the wiretap channel, and $\mathcal{P}_n(W, \delta_n)$ is the index set of δ_n -poor bit-channels, as defined in (50).

Proof. Recall that our encoder first constructs the vector \mathbf{V} with $\mathbf{V}_{\mathcal{A}} = \mathbf{U}$, $\mathbf{V}_{\mathcal{B}} = \mathbf{0}$, and $\mathbf{V}_{\mathcal{R}}$ uniform over $\{0, 1\}^r$. Hence

$$I(\mathbf{U}; \mathbf{Z}) = I(\mathbf{V}_{\mathcal{A}}; \mathbf{Z}) = I(\mathbf{V}_{\mathcal{A} \cup \mathcal{B}}; \mathbf{Z})$$

Since the sets $\mathcal{A}, \mathcal{B}, \mathcal{R}$ partition $[n]$, the vector $\mathbf{V}_{\mathcal{A} \cup \mathcal{B}}$ can be regarded as the input to the induced channel $\mathcal{Q}_n(W, \mathcal{R})$. Moreover, since $\mathcal{R}^c = \mathcal{P}_n(W, \delta_n)$ by (51), Lemma 15 implies that

$$I(\mathbf{V}_{\mathcal{A} \cup \mathcal{B}}; \mathbf{Z}) \leq \mathcal{C}(\mathcal{Q}_n(W, \mathcal{R})) \leq \sum_{i \in \mathcal{P}_n(W, \delta_n)} \mathcal{C}(W_i)$$

The proposition now follows by observing that $\mathcal{C}(W_i) \leq \delta_n$ for all $i \in \mathcal{P}_n(W, \delta_n)$, by the definition of $\mathcal{P}_n(W, \delta_n)$ in (50). ■

Note that we are still free to specify the function δ_n in (51) and (79). This means that the security of our coding scheme is *tunable*. Let us henceforth refer to δ_n as the **security function**; this function is a *design parameter* in our scheme. Proposition 16 implies that choosing different settings for the security function guarantees different levels of security.

Theorem 17. *For any security function such that $\delta_n = o(1/n)$, the strong-security coding scheme guarantees strong security.*

Proof. Follows from Proposition 16, along with the definition of strong security in (3) and the fact that $|\mathcal{P}_n(W, \delta_n)| \leq n$. ■

In fact, we shall see in the next subsection (cf. Theorem 21) that we can achieve the secrecy capacity, while setting the security function to be as small as $\delta_n = 2^{-n^\beta}$ for any positive constant $\beta < 1/2$. In this case, our coding scheme guarantees that the mutual information between the message \mathbf{U} and Eve’s observations \mathbf{Z} scales roughly as

$$I(\mathbf{U}; \mathbf{Z}) = o\left(2^{-\sqrt{n}/n^\epsilon}\right) \quad (80)$$

for any $\epsilon > 0$. Note that this holds regardless of the a priori distribution of the message.

E. Rate of the Strong-Security Coding Scheme

Let $R_n = k/n = |\mathcal{A}|/n$ denote the rate of the strong-security coding scheme, where \mathcal{A} is the set defined in (52). Note that

$$\mathcal{A} = \mathcal{P}_n(W, \delta_n) \setminus \mathcal{B}_n(W^*, \beta)$$

since the sets $\mathcal{G}_n(W^*, \beta)$ and $\mathcal{B}_n(W^*, \beta)$ of good and bad channels in (13), (14) are complements of each other. It follows that

$$R_n \geq \frac{|\mathcal{P}_n(W, \delta_n)|}{n} - \frac{|\mathcal{B}_n(W^*, \beta)|}{n} \quad (81)$$

The asymptotic behavior of the fraction $|\mathcal{B}_n(W^*, \beta)|/n$ is given by Theorem 1. Therefore, in order to prove that the rate of the strong-security coding scheme approaches secrecy capacity, it remains to analyze the asymptotic behavior of $|\mathcal{P}_n(W, \delta_n)|/n$.

In this regard, a recent result of Hassani and Urbanke [16] will be useful. To describe this result, we need to introduce some notation. Given a BSM channel W and a positive $\gamma < 1$, let

$$\mathcal{P}'_n(W, \gamma) \stackrel{\text{def}}{=} \left\{ i \in [n] : Z(W_i) \geq 1 - \gamma \right\} \quad (82)$$

This is similar to the definition of the set $\mathcal{P}_n(W, \delta)$ in (50), except that (82) uses the Bhattacharyya parameters $Z(W_i)$ instead of the channel capacities $\mathcal{C}(W_i)$. Given a positive integer m and a real number ξ in the open interval $(0, 1)$, let $a = a(m, \xi)$ denote the unique positive integer such that

$$\sum_{i=a}^m \binom{m}{i} \leq \xi 2^m < \sum_{i=a-1}^m \binom{m}{i} \quad (83)$$

and define $\alpha(m, \xi) \stackrel{\text{def}}{=} a(m, \xi)/m$. Further, we say that a function $f(m)$ from the positive integers to the interval $(0, 1)$ is in the intersection of $o(1/\sqrt{m})$ and $\omega(1/m)$ if

$$\lim_{m \rightarrow \infty} (\sqrt{m}f(m)) = 0 \quad \text{and} \quad \lim_{m \rightarrow \infty} (mf(m)) = \infty$$

The following is (the second part of) Theorem 3 of Hassani and Urbanke [16]. Although this result is more general than what we need, we state it below exactly as in [16, Theorem 3].

Theorem 18. *Let W be a BSM channel, and let $\xi < 1$ be a positive constant. Fix an arbitrary function $f(m)$ in the intersection of $o(1/\sqrt{m})$ and $\omega(1/m)$, and for all $n = 2^m$ define*

$$\gamma_n \stackrel{\text{def}}{=} 2^{-n^{\alpha(m, \xi)(1+f(m))}} \quad (84)$$

where $\alpha(m, \xi)$ is the function defined in (83). Then the asymptotic behavior of the fraction $|\mathcal{P}'_n(W, \gamma_n)|/n$ is given by

$$\lim_{n \rightarrow \infty} \frac{|\mathcal{P}'_n(W, \gamma_n)|}{n} = \xi(1 - \mathcal{C}(W)) \quad (85)$$

In Corollary 19 and Proposition 20, we specialize the general result of Theorem 18 to our needs.

Corollary 19. *Let W be an arbitrary BSM channel. Then for any positive constant $\beta < 1/2$ we have*

$$\lim_{n \rightarrow \infty} \frac{|\mathcal{P}'_n(W, 2^{-n^\beta})|}{n} = 1 - \mathcal{C}(W) \quad (86)$$

Proof. Applying the Stirling formula to both sides of (83), it can be shown (cf. [16]) that

$$a(m, \xi) = \frac{m}{2} + \frac{Q^{-1}(\xi)}{2} \sqrt{m} + o(\sqrt{m})$$

where $Q(x)$ is the probability that a standard normal random variable will obtain a value larger than x . Consequently, for all positive $\xi < 1$, there exists a positive constant c_ξ such that

$$a(m, \xi) \geq \frac{m}{2} - c_\xi \sqrt{m}$$

for all sufficiently large m . This implies that for any positive constant $\beta < 1/2$, any function $f(m)$ from the positive integers

to the interval $(0, 1)$, and for all sufficiently large m , the following inequality holds

$$\alpha(m, \xi)(1 + f(m)) > \frac{a(m, \xi)}{m} \geq \frac{1}{2} - \frac{c_\xi}{\sqrt{m}} \geq \beta$$

This, in turn, implies that for all sufficiently large $n = 2^m$, we have $\gamma_n < 2^{-n^\beta}$ where γ_n is the function defined in (84). Therefore, the set $\mathcal{P}'_n(W, \gamma_n)$ is a subset of the set $\mathcal{P}'_n(W, 2^{-n^\beta})$ for all sufficiently large n .

Let us assume for a moment that the limit on the left-hand side of (86) exists, call it L . If so, we can conclude from Theorem 18, along with the fact that $\mathcal{P}'_n(W, \gamma_n) \subset \mathcal{P}'_n(W, 2^{-n^\beta})$ for all sufficiently large n , that

$$\lim_{n \rightarrow \infty} \frac{|\mathcal{P}'_n(W, 2^{-n^\beta})|}{n} \geq \lim_{n \rightarrow \infty} \frac{|\mathcal{P}'_n(W, \gamma_n)|}{n} = \xi(1 - \mathcal{C}(W))$$

Since this holds for all positive $\xi < 1$, the lowest possible value of L is $1 - \mathcal{C}(W)$. Now observe that the set $\mathcal{P}'_n(W, 2^{-n^\beta})$ and the set $\mathcal{G}_n(W, \beta)$ of good channels, defined in (13), do not intersect. Therefore, for all $n = 2^m$ we have

$$\frac{|\mathcal{P}'_n(W, 2^{-n^\beta})|}{n} \leq 1 - \frac{|\mathcal{G}_n(W, \beta)|}{n}$$

Together with Theorem 1, this implies that the limit L indeed exists, and is equal to $1 - \mathcal{C}(W)$. ■

Proposition 20. *Let W be an arbitrary BSM channel, and let $\beta < 1/2$ be a positive constant. Further, let $\mathcal{P}_n(W, \delta_n)$ be the index set of δ_n -poor bit-channels, as defined in (50), and suppose that there exist positive constants c_1 and c_2 such that*

$$c_1 2^{-n^\beta} \leq \delta_n \leq 1 - c_2 \quad (87)$$

for all sufficiently large n . Then the asymptotic behavior of the fraction $|\mathcal{P}_n(W, \delta_n)|/n$ is given by

$$\lim_{n \rightarrow \infty} \frac{|\mathcal{P}_n(W, \delta_n)|}{n} = 1 - \mathcal{C}(W) \quad (88)$$

Proof. Let W_1, W_2, \dots, W_n denote the bit-channels, as before. It was shown by Arıkan in [3, Proposition 1] that

$$\mathcal{C}(W_i) \leq \sqrt{1 - Z(W_i)^2} \quad (89)$$

for all $i \in [n]$. Fix a positive constant α such that $\beta < \alpha < 1/2$. Since $Z(W_i) \geq 1 - 2^{-n^\alpha}$ for all $i \in \mathcal{P}'_n(W, 2^{-n^\alpha})$ by definition, Arıkan's bound (89) implies that

$$\mathcal{C}(W_i) \leq 2^{-(n^\alpha - 1)/2} \quad (90)$$

for $i \in \mathcal{P}'_n(W, 2^{-n^\alpha})$. For all sufficiently large n , the right-hand side of (90) is less than the left-hand side of (87), and therefore

$$\mathcal{P}'_n(W, 2^{-n^\alpha}) \subseteq \mathcal{P}_n(W, \delta_n) \quad (91)$$

Also note that the condition $\delta_n \leq 1 - c_2$ on the right-hand side of (87) implies that for all sufficiently large n , we have

$$\mathcal{P}_n(W, \delta_n) \cap \mathcal{G}_n(W, \beta) = \emptyset \quad (92)$$

The proposition now follows by combining (91) and (92) with Corollary 19 and Theorem 1, respectively. ■

We are now ready to prove that for a wide range of security functions, our strong-security coding scheme operates at a rate that approaches the secrecy capacity.

Theorem 21. *For any security function δ_n that satisfies (87), the rate R_n of the corresponding strong-security coding scheme approaches the secrecy capacity, namely*

$$\lim_{n \rightarrow \infty} R_n = \mathcal{C}(W^*) - \mathcal{C}(W) \quad (93)$$

Proof. A lower bound on the rate R_n is given in (81). Along with Proposition 20 and Theorem 1, this immediately shows that $\lim_{n \rightarrow \infty} R_n \geq \mathcal{C}(W^*) - \mathcal{C}(W)$. In order to establish equality in (93), let us partition the set $\mathcal{R} = [n] \setminus \mathcal{P}_n(W, \delta_n)$ in (51) into two subsets, as in Figure 2. These subsets are defined as follows:

$$\mathcal{X} \stackrel{\text{def}}{=} \mathcal{R} \cap \mathcal{B}_n(W^*, \beta) \quad (94)$$

$$\mathcal{Y} \stackrel{\text{def}}{=} \mathcal{R} \cap \mathcal{G}_n(W^*, \beta) \quad (95)$$

Note that the set \mathcal{X} in (94) and the set \mathcal{B} in (53) form a partition of $\mathcal{B}_n(W^*, \beta)$, as illustrated in Figure 2. It follows that

$$R_n = \frac{|\mathcal{P}_n(W, \delta_n)|}{n} - \frac{|\mathcal{B}_n(W^*, \beta)|}{n} + \frac{|\mathcal{X}|}{n} \quad (96)$$

We will show in Proposition 22 of the next subsection that the fraction $|\mathcal{X}|/n$ vanishes as $n \rightarrow \infty$. With this, the theorem follows from (96), Proposition 20, and Theorem 1. ■

F. Reliability of the Strong-Security Coding Scheme

If the main channel W^* is noiseless, Bob can trivially recover the message with probability 1 as follows. Bob receives the vector $\mathbf{Y} = \mathbf{V}G_n$. Since the Arkan transform matrix $G_n = P_n G^{\otimes m}$ is its own inverse over \mathbb{F}_2 , Bob can compute $\mathbf{V} = \mathbf{Y}G_n$ and set $\hat{\mathbf{U}} = \mathbf{V}_A$. Clearly $\hat{\mathbf{U}} = \mathbf{U}$, and condition (1) is trivially satisfied.

What happens if the main channel W^* is not noiseless? Suppose that Bob attempts to use the successive cancellation decoder, as in Section V. Then, according to Theorem 2 and Proposition 3, Bob's probability of error is upper-bounded by the sum of the Bhattacharyya parameters $Z(W_i^*)$ of those bit-channels that are not fixed (to zero). The index set of the bit-channels that are not fixed in our strong-security coding scheme is given by

$$\mathcal{A} \cup \mathcal{R} = \mathcal{G}_n(W^*, \beta) \cup \mathcal{X}$$

where \mathcal{X} is the set defined in (94). The sum of the Bhattacharyya parameters $Z(W_i^*)$ over the set $\mathcal{G}_n(W^*, \beta)$ is bounded by 2^{-n^β} by definition, and therefore

$$\Pr\{\hat{\mathbf{U}} \neq \mathbf{U}\} \leq 2^{-n^\beta} + \sum_{i \in \mathcal{X}} Z(W_i^*) \quad (97)$$

Unfortunately, we are not aware of any useful bounds on the sum $\sum_{i \in \mathcal{X}} Z(W_i^*)$. Thus we do not have a proof that the strong-security coding scheme satisfies the reliability condition (1).

Observe, however, that the security and reliability requirements are fundamentally different. Whether we're interested in the theory or in the practice of wiretap channels, security always requires a proof. It cannot be established through a computational procedure, such as simulation. On the other hand, re-

liability can be (and often is) established through computation in practice. For example, it is very common to rely on simulations to verify the reliability performance of error-correcting codes. We believe that, in practice, our strong-security coding scheme can be decoded to achieve low probabilities of error.

First, the following proposition shows that if W is degraded with respect to W^* , then the set \mathcal{X} that gives us trouble in successive cancellation decoding is small.

Proposition 22. *Let \mathcal{X} be the index set of bit-channels that are not δ_n -poor for Eve yet bad for Bob, as defined in (94). Then*

$$\lim_{n \rightarrow \infty} \frac{|\mathcal{X}|}{n} = 0$$

for any security function δ_n that satisfies (87), provided Eve's channel W is degraded with respect to Bob's channel W^ .*

Proof. We claim that the sets \mathcal{X} , $\mathcal{G}_n(W^*, \beta)$, and $\mathcal{P}_n(W^*, \delta_n)$ are pairwise disjoint, and therefore

$$\frac{|\mathcal{X}|}{n} + \frac{|\mathcal{G}_n(W^*, \beta)|}{n} + \frac{|\mathcal{P}_n(W^*, \delta_n)|}{n} \leq 1 \quad (98)$$

Since $\mathcal{X} \subseteq \mathcal{B}_n(W^*, \beta)$ by definition, and $\mathcal{B}_n(W^*, \beta)$ is the complement of $\mathcal{G}_n(W^*, \beta)$, it is clear that $\mathcal{X} \cap \mathcal{G}_n(W^*, \beta) = \emptyset$. It is also clear that $\mathcal{P}_n(W^*, \delta_n) \cap \mathcal{G}_n(W^*, \beta) = \emptyset$, as in (92). Furthermore, since $\mathcal{X} \subseteq \mathcal{R}$ and $\mathcal{R} = [n] \setminus \mathcal{P}_n(W, \delta_n)$ by definition, we have $\mathcal{X} \cap \mathcal{P}_n(W, \delta_n) = \emptyset$. Consequently, in order to prove our claim in (98), it would suffice to show that

$$\mathcal{P}_n(W^*, \delta_n) \subseteq \mathcal{P}_n(W, \delta_n)$$

But this follows immediately from Lemma 4 along with the definition of the set of δ_n -poor channels in (50). Now observe that as $n \rightarrow \infty$, the fraction $|\mathcal{G}_n(W^*, \beta)|/n$ converges to the capacity $\mathcal{C}(W^*)$ by Theorem 1, whereas the fraction $|\mathcal{P}_n(W^*, \delta_n)|/n$ converges to $1 - \mathcal{C}(W^*)$ by Proposition 20. Together with (98), this completes the proof of the proposition. ■

Depending upon how small the set \mathcal{X} turns out to be in practice, various decoding solutions are potentially applicable.

First, it is quite possible that $\mathcal{X} = \emptyset$ in many situations, especially if the main channel is much better than the wiretap channel. In this case, successive cancellation decoding can be used "as is," and Bob's probability of error is at most 2^{-n^β} .

The following example illustrates this situation. In order to obtain the numerical values given in this example, we have used the methods of [38] to evaluate the polar bit-channels.

Example. Suppose that both the main channel and the wiretap channel are binary symmetric channels, say $C_1 = \text{BSC}(p_1)$ and $C_2 = \text{BSC}(p_2)$. Let us further assume that $p_1 = 10^{-3}$ and that the error-rate required at the output of the main-channel decoder is 10^{-9} . This is often the case in optical fiber communications [32]. We will use the polar transformation (5) of length $n = 2^{20}$. Indeed, codes of this length are already in use today in proprietary 100 GbE fiber-optic systems. We also adopt the following stringent security criterion: we require that the mutual information $I(\mathbf{U}; \mathbf{Z})$ between messages at the input to our encoder and observations at the output of the wiretap channel is less than 10^{-30} . Using the methods developed in this section, we can simultaneously guarantee reliability of 10^{-9} and secu-

urity of 10^{-30} at communication rates close to the secrecy capacity. The following table:

p_2	Rate R	% of C_s
0.45	0.933	95.1%
0.40	0.882	91.9%
0.35	0.817	88.5%
0.30	0.738	84.8%
0.25	0.647	80.9%
0.20	0.543	76.4%
0.15	0.425	71.1%
0.10	0.293	64.0%

summarizes these rates as a function of the bit error-rate p_2 of the wiretap channel. The third column in the table gives the ratio R/C_s , expressed as a percentage. Notably, in *all of these cases*, we have $\mathcal{X} = \emptyset$. In fact, the set \mathcal{X} remains empty until the bit error-rate on the wiretap channel decreases down to $p_2 = 0.066$, in which case $|\mathcal{X}| = 1$. But the secrecy capacity for $p_2 \leq 0.066$ is less than 0.3395, which is probably too small to be of practical interest (in fiber-optic communications). \square

Now suppose that the set \mathcal{X} is nonempty. Observe that this set is fixed and known a priori to all the parties (Alice, Bob, and Eve). In successive cancellation decoding, Bob makes his decisions $\hat{v}_1, \hat{v}_2, \dots, \hat{v}_n$ sequentially using the decision rule (12). Bob will know a priori that this decision rule is unreliable whenever an index $i \in \mathcal{X}$ is reached. Therefore, Bob could follow *both* alternatives $\hat{v}_i = 0$ and $\hat{v}_i = 1$ for all $i \in \mathcal{X}$. Doing so increases the decoding complexity by a factor of $2^{|\mathcal{X}|}$. But if $|\mathcal{X}|$ is a small constant (say, \mathcal{X} contains only a couple of bit-channels), this is not unreasonable.

What can Bob do if the set \mathcal{X} is larger? It is well known that in successive cancellation decoding, a single incorrect decision affects all the following decisions, making them unreliable. We propose to *take advantage* of this phenomenon in order to reduce the decoding complexity. Let i_1 be the smallest index in \mathcal{X} . Suppose that upon branching with $\hat{v}_{i_1} = 0$ and $\hat{v}_{i_1} = 1$, the decoder begins to compute the channel noise (e.g. the Hamming distance to the received vector on a BSC channel) accumulated along each of the two decision paths being followed. Due to the “error propagation” induced by the incorrect decision at i_1 , we expect this estimated noise to accumulate rapidly along the incorrect path. On the other hand, along the correct path, the channel noise should accumulate slowly, governed by the statistics of W^* that are known a priori. This means that the decoder can detect, with high probability, which of the two paths being followed is incorrect. Once the decoder finds that the path with the higher accumulated noise is sufficiently unlikely, according to the channel statistics, this path can be safely discarded.

Of course, it is possible that the second smallest index $i_2 \in \mathcal{X}$ is reached before one of the two paths opened at $i_1 \in \mathcal{X}$ can be discarded. In this case, the decoder would need to begin following four paths. Once the third index $i_3 \in \mathcal{X}$ is reached, the decoder might be following 1, 2, 3, or 4 paths. And so on. In practice, one could design the decoder to follow at most M paths, where M is a pre-determined limit dictated by the decoder complexity considerations. The situation is quite similar to decision-feedback equalization on ISI channels using a bank of M zero-

forcing DFEs. That scenario was analyzed in [43], where it is shown that error-propagation caused by incorrect decisions can be used to discard erroneous decision-feedback paths. It is also shown in [43] that, in practice, small values of M often suffice to achieve very good performance.

We have limited our consideration herein to successive cancellation decoding, or variants thereof. It is also possible that other methods of decoding polar codes (such as belief propagation [18] or recursive-list decoding [13]) may be relatively robust to not fixing a small set \mathcal{X} of bad channels. Analysis of such decoders is a research problem of independent interest.

VII. DISCUSSION AND OPEN PROBLEMS

We briefly mention certain straightforward extensions of our results. So far, we have considered exclusively binary-input symmetric wiretap channels. However, it is well known that the polarization phenomenon extends to other types of channels. Arkan shows in [3] that, given a non-symmetric binary-input channel W , polar codes achieve its symmetric capacity $\mathcal{I}(W)$ in the *average sense*. This implies that our coding scheme achieves the symmetric capacity difference $\mathcal{I}(W^*) - \mathcal{I}(W)$, also in the average sense. Specifically, suppose we modify our encoding algorithm in Section IV as follows. Instead of constructing the vector $\mathbf{v} \in \{0, 1\}^n$ by setting $\mathbf{v}_R = \mathbf{e}$, $\mathbf{v}_A = \mathbf{u}$, and $\mathbf{v}_B = \mathbf{0}$, we set $\mathbf{v}_R = \mathbf{e}$, $\mathbf{v}_A = \mathbf{u}$, and $\mathbf{v}_B = \mathbf{s}$, where \mathbf{s} is a fixed binary vector known a priori to all the parties. Then there exists *some choice* of \mathbf{s} such that Theorem 8 and Theorem 9 hold. Based upon the results of Arkan in [3], exactly the same proof as before (Lemma 6 and Lemma 7) applies.

Our results also extend to discrete memoryless channels with non-binary input. It was recently proved in [35] that channels with an input alphabet of *prime size* q are polarized by the same transformation (5), and the corresponding versions of Theorem 1 and Theorem 2 hold. The probability of error under successive cancellation decoding still scales as $O(2^{-n^\beta})$ for all prime q . This means that our proof of Lemmas 6 and 7 goes through essentially “as is” (as long as r is replaced by $r \log_2 q$ throughout). If the size q of the input alphabet is not prime, polarization requires either a randomized permutation on the input or multilevel coding (see [35] for more details). It can be shown that our results in Section V extend to this case as well.

It is not clear whether the strong-security results of the previous section can be similarly extended to non-symmetric and/or to non-binary-input wiretap channels. We believe they can, and pose a proof of this as an open problem.

Another open problem of great interest is how to code for the situation where the wiretap channel W is *not* degraded with respect to the main channel W^* . Note that channel degradation is sufficient but, to the best of our knowledge, not necessary for our coding scheme to work. What seems to be necessary is that the set of bit-channels that are “good” for Eve but “bad” for Bob is either empty (as in Section V) or at least very small (as in Section VI). Unfortunately, in the general case, there is no reason why the number of such bit-channels could not be large.

Finally, we point out that all the constructions in this paper are only as explicit as the polar codes themselves. An exact algorithm for computing the sets $\mathcal{G}_n(W, \beta)$ and $\mathcal{P}_n(W, \delta)$ in (13)

and (50) was given by Arıkan in [3]. However, this algorithm requires time and memory that grow exponentially with the code length n . Since then, several heuristic algorithms for this problem have been proposed [2, 30, 31]. However, these algorithms do not provide useful guarantees on the quality of their output. Such guarantees are clearly essential to establish the security of our coding scheme. Fortunately, the problem has been resolved in [38]. The algorithm of [38] runs in linear time, and makes it possible to compute upper and lower bounds on the capacity of polar bit-channels with an arbitrary degree of precision. For example, to establish the results reported in (99), we have run the algorithm of [38] with a precision of 300 bits (which is necessary to provide meaningful guarantees of security down to 10^{-30}).

ACKNOWLEDGMENT

We thank Mihir Bellare, Hamed Hassani, Satish Korada, Ryutaroh Matsumoto, Ido Tal, Emre Telatar, and Rüdiger Urbanke for very helpful discussions. We are especially grateful to Satish Korada for pointing us to Lemma 4.7 of his Ph.D. thesis, and to Ido Tal for his help with the example in Section VI-F.

REFERENCES

- [1] M. Andersson, V. Rathi, R. Thobaben, J. Kliewer, and M. Skoglund, "Nested polar codes for wiretap and relay channels," preprint of June 17, 2010, arxiv.org/abs/1006.3573.
- [2] E. Arıkan, "A performance comparison of polar codes and Reed-Muller codes," *IEEE Comm. Letters*, vol. 12, no. 6, pp. 447–449, June 2008.
- [3] E. Arıkan, "Channel polarization: A method for constructing capacity-achieving codes for symmetric binary-input memoryless channels," *IEEE Trans. Inform. Theory*, vol. 55, no. 7, pp. 3051–3073, July 2009.
- [4] E. Arıkan and E. Telatar, "On the rate of channel polarization," preprint of July 24, 2008, arxiv.org/abs/0807.3806.
- [5] M. Bellare, S. Tessaro, and A. Vardy, *A cryptographic treatment of the wiretap channel*, preprint, April 2010.
- [6] M. Bellare, A. Desai, E. Jorjipii, and Ph. Rogaway, "A concrete security treatment of symmetric encryption," *Proc. IEEE Symp. Foundations Computer Science*, pp. 394–403, Miami Beach, FL., October 1997.
- [7] C. Bennett, G. Brassard, C. Crépeau, and U.M. Maurer, "Generalized privacy amplification," *IEEE Trans. Inform. Theory*, vol. 41, no. 6, pp. 1915–1923, November 1995.
- [8] M. Cheraghchi, F. Didier, and A. Shokrollahi, "Invertible extractors and wiretap protocols," preprint of January 14, 2009, arxiv.org/abs/0901.2120.
- [9] G. Cohen and G. Zémor, "The wire-tap channel applied to biometrics," *Proc. Intern. Symp. Information Theory and its Applications*, Parma, Italy, October 2004.
- [10] G. Cohen and G. Zémor, "Syndrome-coding for the wiretap channel revisited," *Proc. IEEE Information Theory Workshop*, pp. 33–36, Chengdu, China, October 2006.
- [11] T.M. Cover and J.A. Thomas, *Elements of Information Theory*. 2nd Ed., Hoboken, NJ: John Wiley & Sons, 2006.
- [12] I. Csiszár and J. Körner, "Broadcast channels with confidential messages," *IEEE Trans. Inform. Theory*, vol. 24, no. 3, pp. 339–348, May 1978.
- [13] I. Dumer and K. Shabunov, "Soft-decision decoding of Reed-Muller codes: recursive lists," *IEEE Trans. Inform. Theory*, vol. 52, no. 3, pp. 1260–1266, March 2006.
- [14] R.G. Gallager, *Information Theory and Reliable Communication*. Hoboken, NJ: John Wiley & Sons, 1968.
- [15] S. Goldwasser and S. Micali, "Probabilistic encryption," *J. Comput. Syst. Sci.*, vol. 28, pp. 270–299, April 1984.
- [16] S. Hassani and R.L. Urbanke, "On the scaling of polar codes: The behavior of polarized channels," *Proc. IEEE Intern. Symp. Information Theory*, pp. 874–878, Austin, TX., June 2010.
- [17] E. Hof and S. Shamai, "Secrecy-achieving polar-coding for binary-input memoryless symmetric wire-tap channels," *IEEE Trans. Inform. Theory*, submitted for publication, May 2010.
- [18] N. Hussami, R.L. Urbanke, and S.B. Korada, "Performance of polar codes for channel and source coding," *Proc. IEEE Intern. Symp. Information Theory*, pp. 1488–1492, Seoul, Korea, June 2009.
- [19] S.B. Korada, *Polar Codes for Channel and Source Coding*, Ph.D. dissertation, EPFL, Lausanne, Switzerland, May 2009.
- [20] S. B. Korada, E. Şaşıoğlu, and R.L. Urbanke, "Polar codes: Characterization of exponent, bounds, and constructions," *Proc. IEEE Intern. Symp. Information Theory*, pp. 1483–1487, Seoul, Korea, June 2009.
- [21] O.O. Koyluoğlu and H. El Gamal, "Polar coding for secure transmission and key agreement," *Proc. IEEE Intern. Symp. Personal Indoor and Mobile Radio Comm.*, pp. 2698–2703, Istanbul, Turkey, September 2010.
- [22] S. Leung-Yan-Cheong, "On a special class of wire-tap channels," *IEEE Trans. Inform. Theory*, vol. 23, no. 5, pp. 625–627, September 1977.
- [23] S. Leung-Yan-Cheong and M.E. Hellman, "The Gaussian wire-tap channel," *IEEE Trans. Inform. Theory*, vol. 24, no. 4, pp. 451–456, July 1978.
- [24] Y. Liang, H.V. Poor, and S. Shamai, "Information theoretic security," *Foundations and Trends in Communications and Information Theory*, vol. 5, issue 4–5, pp. 355–580, 2008.
- [25] H. Mahdaviyar and A. Vardy, "Achieving the secrecy capacity of wiretap channels using polar codes," preprint of December 31, 2009, arxiv.org/abs/1001.0210.
- [26] H. Mahdaviyar and A. Vardy, "Achieving the secrecy capacity of wiretap channels using polar codes," *Proc. IEEE Intern. Symp. Information Theory*, pp. 913–917, Austin, TX., June 2010.
- [27] R. Matsumoto, "On the randomness in the encoder," private communication, June 2010.
- [28] U.M. Maurer, "The strong secret key rate of discrete random triples," pp. 271–285 in *Communication and Cryptography — Two Sides of One Tapestry*, R.E. Blahut et al. (Eds.), Boston: Kluwer Academic, 1994.
- [29] U.M. Maurer and S. Wolf, "Information-theoretic key agreement: From weak to strong secrecy for free," *Lect. Notes Computer Science*, vol. 1807, pp. 351–368, Springer-Verlag, 2000.
- [30] R. Mori and T. Tanaka, "Performance and construction of polar codes on symmetric binary-input memoryless channels," *Proc. IEEE Intern. Symp. Information Theory*, pp. 1496–1500, Seoul, Korea, June 2009.
- [31] R. Mori and T. Tanaka, "Performance of polar codes with the construction using density evolution," *IEEE Comm. Letters*, vol. 13, no. 7, pp. 519–521, July 2009.
- [32] D.K. Mynbaev and L.L. Scheiner, *Fiber-Optic Communications Technology*. Englewood Cliffs, NJ: Prentice Hall, 2000.
- [33] L.H. Ozarow and A.D. Wyner, "Wire-tap channel II," *Bell Lab. Tech. J.*, vol. 63, no. 10, pp. 2135–2157, December 1984.
- [34] Y. Polyanskiy, H.V. Poor, and S. Verdú, "Channel coding rate in the finite blocklength regime," *IEEE Trans. Inform. Theory*, vol. 56, no. 5, pp. 2307–2359, May 2010.
- [35] E. Şaşıoğlu, E. Telatar, and E. Arıkan, "Polarization for arbitrary discrete memoryless channels," *Proc. IEEE Information Theory Workshop*, pp. 144–148, Taormina, Italy, October 2009.
- [36] V. Strassen, "Asymptotische Abschätzungen in Shannon's Informations-theorie," *Trans. 3-rd Prague Conf. Information Theory, Statistical Decision Functions, and Random Processes*, pp. 689–732, Publ. Czech. Acad. Sci., Prague, 1964.
- [37] A. Suresh, A. Subramanian, A. Thangaraj, M. Bloch, and S.W. McLaughlin, "Strong secrecy for erasure wiretap channels," *Proc. IEEE Information Theory Workshop*, Dublin, Ireland, September 2010.
- [38] I. Tal and A. Vardy, "How to construct polar codes," presented at the *IEEE Information Theory Workshop*, Dublin, Ireland, September 2010.
- [39] A. Thangaraj, S. Dihidar, A.R. Calderbank, S.W. McLaughlin, and J. Merolla, "Applications of LDPC codes to the wiretap channel," *IEEE Trans. Inform. Theory*, vol. 53, no. 8, pp. 2933–2945, August 2007.
- [40] M. Van Dijk, "On a special class of broadcast channels with confidential messages," *IEEE Trans. Inform. Theory*, vol. 43, no. 2, pp. 712–714, March 1997.
- [41] V.K. Wei, "Generalized Hamming weights for linear codes," *IEEE Trans. Inform. Theory*, vol. 37, no. 5, pp. 1412–1418, September 1991.
- [42] A.D. Wyner, "The wire-tap channel," *Bell System Tech. J.*, vol. 54, no. 8, pp. 1355–1387, October 1975.
- [43] D. Yellin, A. Vardy, and O. Amrani, "Joint equalization and coding for intersymbol interference channels," *IEEE Trans. Inform. Theory*, vol. 43, no. 2, pp. 409–425, March 1997.

Hessam MahdaviFar (S'10) received the B.Sc. degree from the Sharif University of Technology, Tehran, Iran, in 2007, and the M.Sc. degree from the University of California-San Diego, La Jolla, in 2009, both in electrical engineering, where he is currently pursuing the Ph.D. degree.

His research interests are in the areas of algebraic coding theory and information theory.

Alexander Vardy (S'88–M'91–SM'94–F'98) was born in Moscow, U.S.S.R., in 1963. He earned his B.Sc. (*summa cum laude*) from the Technion, Israel, in 1985, and Ph.D. from the Tel-Aviv University, Israel, in 1991.

During 1985–1990 he was with the Israeli Air Force, where he worked on electronic counter measures systems and algorithms. During the years 1992 and 1993 he was a Visiting Scientist at the IBM Almaden Research Center, in San Jose, CA. From 1993 to 1998, he was with the University of Illinois at Urbana-Champaign, first as an Assistant Professor then as an Associate Professor. He is now a Professor in the Department of Electrical Engineering, the Department of Computer Science, and the Department of Mathematics, all at the University of California San Diego (UCSD). While on sabbatical from UCSD, he has held long-term visiting appointments with CNRS, France, the EPFL, Switzerland, and the Technion, Israel. His research interests include error-correcting codes, algebraic and iterative decoding algorithms, lattices and sphere packings, coding for digital media, cryptography and computational complexity theory, and fun math problems.

Prof. Vardy received an IBM Invention Achievement Award in 1993, and NSF Research Initiation and CAREER awards in 1994 and 1995. In 1996, he was appointed Fellow in the Center for Advanced Study at the University of Illinois, and received the Xerox Award for faculty research. In the same year, he became a Fellow of the Packard Foundation. He received the IEEE Information Theory Society Paper Award (jointly with Ralf Koetter) for the year 2004. In 2005, he received the Fulbright Senior Scholar Fellowship, and the Best Paper Award at the IEEE Symposium on Foundations of Computer Science (FOCS). During 1995–1998, he was an Associate Editor for Coding Theory and during 1998–2001, he was the Editor-in-Chief of the IEEE TRANSACTIONS ON INFORMATION THEORY. He was also an Editor for the *SIAM Journal on Discrete Mathematics*. He has been a member of the Board of Governors of the IEEE Information Theory Society from 1998 to 2006, and again starting in 2011.